

## Durham Research Online

---

### Deposited in DRO:

21 August 2020

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Wang, Yuchan and Li, Baojiu and Cautun, Marius (2020) 'Iterative removal of redshift space distortions from galaxy clustering.', *Monthly notices of the Royal Astronomical Society*, 497 (3). pp. 3451-3471.

### Further information on publisher's website:

<https://doi.org/10.1093/mnras/staa2136>

### Publisher's copyright statement:

This article has been accepted for publication in *Monthly notices of the Royal Astronomical Society*. ©: 2020 The Author(s). Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

### Additional information:

---

## Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Iterative removal of redshift-space distortions from galaxy clustering

Yuchan Wang,<sup>1</sup>★ Baojiu Li<sup>2</sup>★ and Marius Cautun<sup>3</sup>★

<sup>1</sup>*Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>2</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>3</sup>*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

Accepted 2020 July 10. Received 2020 July 8; in original form 2019 December 6

## ABSTRACT

Observations of galaxy clustering are made in redshift space, which results in distortions to the underlying isotropic distribution of galaxies. These redshift-space distortions (RSDs) not only degrade important features of the matter density field, such as the baryonic acoustic oscillation (BAO) peaks, but also pose challenges for the theoretical modelling of observational probes. Here, we introduce an iterative non-linear reconstruction algorithm to remove RSD effects from galaxy clustering measurements, and assess its performance by using mock galaxy catalogues. The new method is found to be able to recover the real-space galaxy correlation function with an accuracy of  $\sim 1$  per cent, and restore the quadrupole accurately to 0, on scales  $s \gtrsim 20 h^{-1}$  Mpc. It also leads to an improvement in the reconstruction of the initial density field, which could help to accurately locate the BAO peaks. An ‘internal calibration’ scheme is proposed to determine the values of cosmological parameters, as a part of the reconstruction process, and possibilities to break parameter degeneracies are discussed. RSD reconstruction can offer a potential way to simultaneously extract the cosmological parameters, initial density field, real-space galaxy positions, and large-scale peculiar velocity field (of the real Universe), making it an alternative to standard perturbative approaches in galaxy clustering analysis, bypassing the need for RSD modelling.

**Key words:** methods: numerical – Galaxy: evolution – dark matter – distance scale – large-scale structure of Universe.

## 1 INTRODUCTION

The observed large-scale cosmic structures today encode information about the primordial matter density field – the earliest memory of our own Universe, that came from a time when the Universe was in a simpler form, where density perturbations can be described by linear perturbation theory and the non-linear structure formation had not made the picture more complicated. As an example, the nearly Gaussian curvature fluctuations, as supported by observations (Ade et al. 2014, 2016; Planck Collaboration et al. 2019), can teach us a lot about what has happened during inflation. The observed Universe today, however, can look very different from its initial conditions, due largely to the growth of tiny density perturbations by gravitational instability to form large, non-linear, dark matter clumps in which galaxies, stars, and planets evolve. Inevitable in this process is the permanent loss of certain details of the primordial state of the Universe, but it is still possible to retrieve the remaining useful information by ‘reconstructing’ the initial condition. The latter is a topic which has been investigated for several decades, with increasing interest in recent years (see e.g. Peebles 1989; Croft & Gaztanaga 1997; Brenier et al. 2003; Eisenstein et al. 2005; Zhu et al. 2017; Zhu, Yu & Pen 2018; Schmittfull, Baldauf & Zaldarriaga 2017; Shi, Cautun & Li 2018; Hada & Eisenstein 2019, 2018; Birkin et al. 2019; Bos, Kitaura & van de Weygaert 2019; Wang & Pen 2019; Yu & Zhu 2019; Zhu et al. 2019; Kitaura et al. 2019; and references therein).

One of the main motivations of initial density reconstruction is related to the extraction of the baryonic acoustic oscillation (BAO) signal from galaxy surveys. BAO is a cosmological relic of the random density fluctuations that propagated in the primordial photon-electron-nuclei plasma before recombination. At the epoch of recombination, the disappearance of free electrons stopped this propagation, so that the perturbations and their interference were frozen, leaving an imprint in the matter distribution that is detectable at late times in the galaxy distribution (Eisenstein & Hu 1998).

This imprint is a typical length scale corresponding to the sound horizon, the largest distance sound waves in the plasma could have travelled by a given time, at recombination. For this reason, BAO serves as valuable standard ruler that can be used to study the expansion history of the Universe.

Precise measurements of cosmological distances using BAO can improve the prospective of constraining cosmological models and shedding light on the mystery of the cosmic acceleration (Weinberg et al. 2013), with forthcoming galaxy surveys (Johnston et al. 2008; DESI Collaboration et al. 2016; Laureijs et al. 2011).

However, the BAO peaks found through the observed galaxy correlation function and power spectrum are shifted, weakened and broadened (Eisenstein et al. 2007; Crocce & Scoccimarro 2008) by the process of non-linear gravitational evolution and bulk motions of matter (Obuljen et al. 2017), making it harder to accurately determine the peak positions and to use them to measure cosmological distances. This is further complicated by the fact that galaxies are biased tracers of the large-scale structure, and by redshift space distortions (RSDs), a phenomenon that arises because we measure the redshifts, rather than real distances, of galaxies, and the former

\* E-mail: yuchan.wang@durham.ac.uk (YW); baojiu.li@durham.ac.uk (BL); marius.cautun@gmail.com (MC)

can be affected by the large-scale peculiar velocity field, leading to incorrect interpreted galaxy coordinates. Both of the latter effects can further degrade the potential of BAO as a standard ruler (Zhu et al. 2017; Birkin et al. 2019). The idea is that with reconstruction we can at least partially remove these effects, therefore improving the accuracy of cosmological constraints.

A variety of previous reconstruction methods have found success in reducing of the effects of cosmic structure formation in the recovery of the BAO peaks. Starting from the first attempt (which is now called standard reconstruction) reversing the motion of galaxies (Eisenstein et al. 2007), which has been proved to be effective in observations (Padmanabhan et al. 2012), improvement has been found in methods using iterations (Schmittfull et al. 2017). Inspired by Lagrangian perturbation theory, which uniquely maps the final Eulerian coordinates of galaxies to a set of initial Lagrangian positions, recent developments propose that the process of reconstruction can be treated as solving an optimal mass assignment problem (Frisch et al. 2002; Brenier et al. 2003; Mohayaee et al. 2003). This problem has been lately solved as a non-linear partial differential equation (PDE) using different algorithms (Zhu et al. 2017; Shi et al. 2018). Forward-modelling reconstruction methods are also studied extensively (e.g. Kitaura & Enßlin 2008; Jasche & Wandelt 2013; Wang et al. 2014; Lavaux 2016), where efficient Monte Carlo samples of the initial density field phases are combined with non-linear evolution to select the initial condition that would match well late-time observations of the local Universe.

The reconstruction method proposed by Shi et al. (2018) is the starting point of the iterative reconstruction scheme to be described in this work. This method reduces the reconstruction problem into solving a Monge-Ampere-type PDE, which gives the mapping between the initial, Lagrangian, and final, Eulerian, coordinates of particles. In three spatial dimensions, the PDE contains up to cubic powers of second-order derivatives, and can be solved using a slightly modified multigrid relaxation technique. Although originally developed for reconstructions from a dark matter field, its generalization for reconstructions from biased tracers, such as galaxies and dark matter haloes, turned out to be straightforward (Birkin et al. 2019). In this work, we will further extend this method for reconstructions from biased tracers in *redshift* space, by making use of the relation between the displacement field and the peculiar velocity field.

As mentioned above, RSD means that the inferred galaxy coordinate is different from its true coordinate. There are two regimes of the RSD effect, as can be illustrated by considering two galaxies, both along the line of sight (LOS), one in front of and the other behind a galaxy cluster which is along the same LOS. If these galaxies are distant from the central cluster, they fall towards the latter but the infall velocity is generally not very high – the galaxy in front of the cluster experiences an additional redshift due to the infall velocity, making it appear further away from us, whilst the one behind has an additional blueshift that makes it appear to be closer to observer than its true distance. In this regime, the two galaxies would appear closer to each other, leading to a squashing (Kaiser) effect along the LOS in the galaxy correlation function. On the other hand, if the two galaxies are both much closer to the cluster centre, their velocities are likely much larger; the one in front could appear to be behind the cluster and vice version, which causes a strongly elongated feature along the LOS in the galaxy correlation function, known as the finger-of-God (FoG) effect. The large-scale Kaiser effect can be well described by linear perturbation theory, whilst the FoG effect, being on small scales, is non-linear. The FoG effect causes ‘trajectory crossing’, i.e. it changes the ranking order of the distances of galaxies, and in

general this poses a limitation on reconstruction as we will discuss later.

Assuming no trajectory crossing and a curl-free velocity field, the peculiar velocity field induced gravitationally by overdensities can be derived from the density field itself in the real space. Intuitively, the RSD effect can be described as a ‘more evolved matter field’ (Taylor & Rowan-Robinson 1993), recognizing this intimate relationship between RSD and gravitational process. Accordingly, in the standard reconstruction approach, the RSD effect has been considered as an additional linear factor on the displacements of galaxies following Kaiser’s equation that links the displacement field to the compression effect due to galaxy coherent motion (but neglects the FoG effect).

Obtaining velocity field from density field through non-linear reconstruction has been explored by Yu & Zhu (2019). Their result suggests that the correlation between the matter density field and the velocity field can be more complicated than the linear theory prediction. Since the non-linear displacement field can be obtained from new reconstruction methods, including that of Shi et al. (2018), we are interested to infer the peculiar velocity from it and subsequently use this information to ‘undo’ the RSD effect on measured galaxy coordinates.

However, estimating velocities from a density field in redshift space is an inverse problem – no real-space density field is known a priori in practice. A reliable way to approach the problem could be to use an iterative approach similar to self-calibration between the real- and redshift-space density fields until one obtains a converged result. It was proposed by Yahil et al. (1991) and Strauss (1989) that an iteration scheme can be used to recover the density field in real space from observations. In the linear regime,  $N$ -body simulation results confirmed the potential of this method (Davis, Strauss & Yahil 1991). However, non-linear effect caused by the random motions of galaxies can lead to erroneous estimations, especially in high-density clusters. This can be mitigated by a smoothing of the velocity field, echoing the result found by Cole, Fisher & Weinberg (1994) where the smoothed field gave a significantly more accurate estimation of redshift distortion parameter,  $\beta$ . A second-order improvement of the method was proposed by Gramann, Cen & Gott (1994), and a quasi-non-linear treatment by Taylor & Rowan-Robinson (1993). They both found a strong correlation with the true density field of the density reconstructed from redshift space. More recently, iterative constructions of the initial density field have been proposed by Hada & Eisenstein (2018), Hada & Eisenstein (2019), extending the work of Monaco & Efstathiou (1999), and by Burden, Percival & Howlett (2015). Our approach follows a similar iterative procedure as these more recent works, but has a number of differences. For examples, instead of reconstructing the initial density field, we aim to reconstruct the galaxy coordinates in real space because we are more interested in the removal of RSD effects from real observational data; our displacement field is obtained from non-linear reconstruction; and we have defined different estimators (mainly in configuration space) to quantitatively examine the reconstruction results during iterations. We are interested in reconstruction in an *internal-calibration* sense, namely the physical and technical parameters used for reconstruction are tuned by inspecting the reconstruction outcome itself.

The paper is organized in the following way. In Section 2, we describe our methodology. In Section 2.1, we introduce the basics of the reconstruction method proposed in Shi et al. (2018). In Section 2.2, we relate the displacement field to the peculiar velocity field, arguing that this link enables an iterative method in which, starting with some rough initial guess of these fields, we can gradually

**Table 1.** A short summary of the symbols used throughout this paper. The first block (from  $\mathbf{x}$  to  $K$ ) contains the various quantities used in the reconstruction process, the second block ( $f$  to  $b_{\text{sim}}$ ) are physical parameters related to the galaxy catalogues, the third block ( $n_g$  to  $C_{\text{ani}}$ ) are technical parameters used in the reconstruction, and the last block [E1 to  $R'(s)$ ] are estimators defined to check the convergence of reconstruction. The first column contains the symbols, the second column their physical meaning and the last column the default values (a ‘—’ is used for quantities without default values). We find that for estimators E1 and E2 the number of iterations required before convergence is generally smaller than for estimators E3 and E4, and so a range of values is given for  $K$ .

Symbol	Physical meaning	Value
$\mathbf{x}$	Real-space galaxy coordinate	—
$r$	Real-space distance	—
$\mathbf{s}$	Redshift-space galaxy coordinate	—
$s$	Redshift-space distance	—
$\mathbf{q}$	Initial (Lagrangian) coordinate	—
$\mathbf{x}^{(k)}$	Reconstructed real-space galaxy coordinate ( $k$ th iteration)	—
$\Psi_S^{(k)}$	Displacement field from reconstruction on smoothed galaxy density field ( $k$ th iteration)	—
$\Psi^{(k)}$	Displacement field from reconstruction on un-smoothed galaxy density field ( $k$ th iteration)	—
$\delta_{\text{ini}}$	Initial matter density field	—
$\delta_g^r$	Final real-space galaxy density field	—
$\delta_g^s$	Final redshift-space galaxy density field	—
$\delta_{\text{rec}}$	Reconstructed matter density field from final real-space galaxy catalogue	—
$\delta_{\text{rec}}^{(k)}$	Reconstructed matter density field from reconstructed real-space galaxy catalogue ( $k$ th iteration)	—
$\delta_g^{(k)}$	Galaxy density field of reconstructed real-space galaxy catalogue ( $k$ th iteration)	—
$\delta_{g,S}^{(k)}$	Smoothed galaxy density field of reconstructed real-space galaxy catalogue ( $k$ th iteration)	—
$r[a, b]$	Cross-correlation coefficients between fields $a$ and $b$	—
$\xi_{\text{gg}}(r)$	Real-space galaxy autocorrelation function	—
$\xi_{\text{gm}}(r)$	Real-space galaxy-matter cross-correlation function	—
$\xi_{\text{gg}}^s(s)$	Redshift-space galaxy autocorrelation function	—
$\xi_{0,2,4}(s)$	Redshift-space galaxy correlation function monopole, quadrupole and hexadecapole	—
$K$	Value of iteration number $k$ at convergence	3–6
$f$	Linear growth rate	0.735
$b^{(k)}$	Linear galaxy bias ( $k$ th iteration)	—
$b_{\text{sim}}$	Linear galaxy bias measured in simulation	1.95
$n_g$	Galaxy number density	$3.2 \times 10^{-4} [h^{-1} \text{Mpc}]^{-3}$
$dx$	Reconstruction grid cell size	$2 h^{-1} \text{Mpc}$
$S$	Isotropic Gaussian smoothing scale	$9 h^{-1} \text{Mpc}$
$C_{\text{ani}}$	Anisotropic smoothing parameter	1.0
E1	$r[\delta_{\text{rec}}^{(k)}, \delta_{\text{ini}}]$	—
E2	$r[\delta_g^{(k)}, \delta_g^r]$	—
E3	$\xi_2[\mathbf{x}^{(k)}](s)$	—
E4	$\xi_0[\mathbf{x}^{(k)}](s)/\xi_{\text{gg}}(r)$	—
$R(s)$	$\xi_0(s)/\xi_0[\mathbf{x}^{(k)}](s)$	—
$R'(s)$	$\xi_0(s)/\xi_{\text{gg}}(r)$	—

improve our knowledge of them during each iteration. In Section 2.4, we describe in great details how the method is implemented in practice and define four estimators to assess its performance. Because of the large number of symbols used in this paper, we summarize them in Table 1 to aim the reader. In Section 3, we test the effect of choosing different physical and technical parameters in our pipeline on the reconstruction result and performance; this section is technical and readers who are more interested in the results can skip it. Section 4 is the main result of this paper, where we show an application of the new method, in which we use mock galaxy catalogues constructed from a suite of  $N$ -body simulations to assess the potential of using this method to simultaneously obtain the real-space galaxy coordinates, the real-space initial matter density field and determine the physical parameters of the cosmological model. Finally, we summarize the main results, discuss the outlook and future applications of the method, and conclude in Section 5.

The main figures of this paper are Fig. 1 (schematic description of the method) and Figs 11 and 12 (performance illustration).

## 2 METHODOLOGY

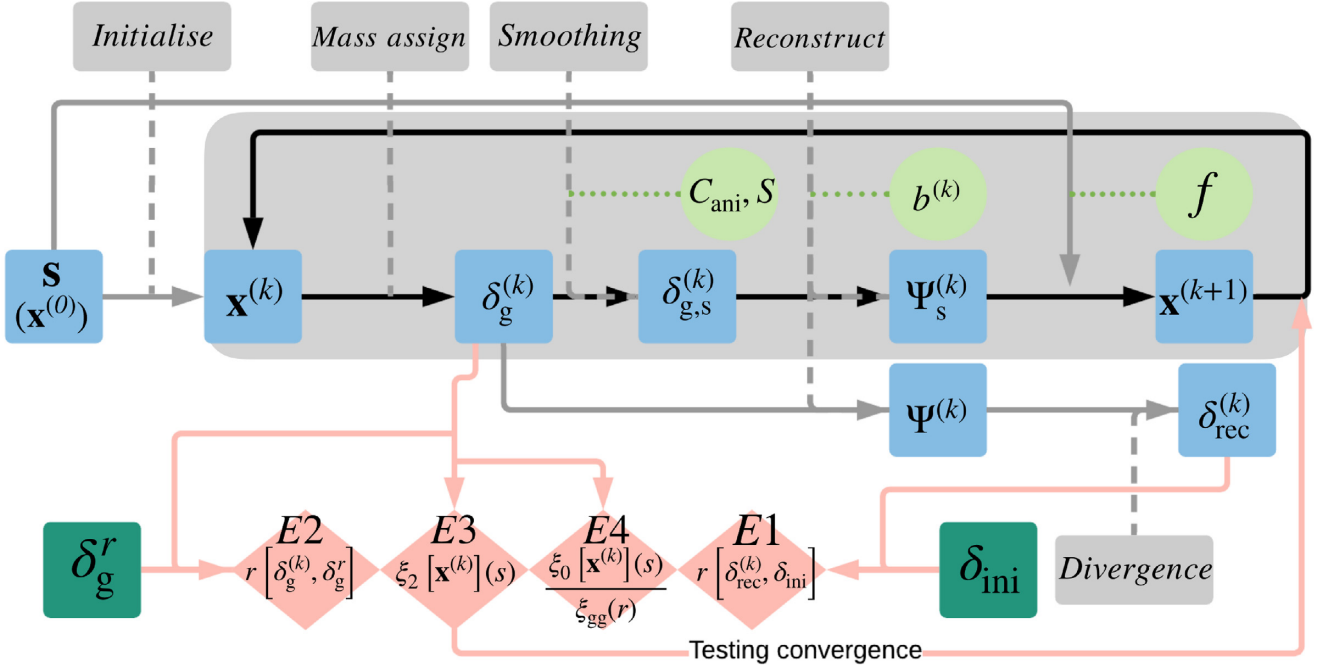
### 2.1 Non-linear reconstruction in real space

The iterative RSD reconstruction method described in this paper is based on the real-space non-linear reconstruction method introduced by Shi et al. (2018); see also Li (2018). For completeness, here we briefly recap the basic idea behind that method.

Our main objective is to identify a mapping between the initial Lagrangian coordinate,  $\mathbf{q}$ , of a particle and its Eulerian coordinate,  $\mathbf{x}(t)$ , at some later time  $t$ . Such a mapping can be uniquely obtained, at least under the condition that the trajectories of particles have not crossed each other, by starting from the following equation:

$$\rho(\mathbf{x})d^3\mathbf{x} = \rho(\mathbf{q})d^3\mathbf{q} \approx \bar{\rho}d^3\mathbf{q}, \quad (1)$$

which is based on continuity equation stating that mass is conserved in an infinitesimal volume element.  $\rho(\mathbf{q})$  and  $\rho(\mathbf{x})$  are, respectively, the initial density field and the density field at time  $t$ . As the density



**Figure 1.** The flowchart indicating the different steps of the iterative reconstruction pipeline introduced in this paper. The light blue boxes (in the third and fourth rows) are the physical quantities as input, intermediate result or output of the pipeline; the grey boxes (in the first and last rows) are operations that take these inputs to produce intermediate results or outputs; the pink diamonds (in the last row) are the estimators defined to assess the performance of reconstruction, and the dark green boxes (in the last row) are the real-density fields that are used for evaluating two of these estimators (E1 and E2); the pink lines with arrows (which link boxes to diamonds) show which quantities are needed to evaluate each estimator; the light green circles indicate the parameters used in the process, which need to be tested and optimized, as we will see in the next section, and the dotted green lines indicate in which operations are these parameters used. See the main text for more details.

field is very close to homogeneous at early times, we can approximate the initial  $\rho(\mathbf{q})$  as a constant,  $\rho(\mathbf{q}) \simeq \bar{\rho}$ .

The displacement field,  $\Psi(\mathbf{x}) = \mathbf{x} - \mathbf{q}$ , between the final and initial positions of a particle can be rewritten as

$$\nabla_{\mathbf{x}} \Theta(\mathbf{x}) \equiv \mathbf{q} = \mathbf{x} - \Psi(\mathbf{x}), \quad (2)$$

where  $\Theta(\mathbf{x})$  is the displacement potential, whose gradient is  $\mathbf{q}$ . Underlying these definitions is another approximation in this method, namely the displacement field is curl-free,  $\nabla \times \Psi = 0$ , which should break down on small scales. Substituting equation (2) into equation (1), we get

$$\det[\nabla^i \nabla_j \Theta(\mathbf{x})] = \frac{\rho(\mathbf{x})}{\bar{\rho}} \equiv 1 + \delta(\mathbf{x}), \quad (3)$$

where  $i, j$  runs over 1, 2, 3, and  $\delta(\mathbf{x})$  is the density contrast at time  $t$ . The symbol ‘det’ denotes the determinant of a matrix, in this case the Hessian of  $\Theta(\mathbf{x})$ . A new algorithm to solve equation (3) was developed in Shi et al. (2018), which reduces the problem into the numerical solution for a non-linear PDE that contains up to the third (in 3D) power of the second-order derivatives of  $\Theta$ . It was later generalized by Birkin et al. (2019) to more generic cases where  $\delta(\mathbf{x})$  in equation (3) is a biased description of the true underlying matter density field. As this work does not extend the numerical algorithm to solve this PDE, we shall omit the technical details here and refer interested readers to those references.

Once  $\Theta(\mathbf{x})$  and therefore  $\Psi(\mathbf{x})$  are obtained, the reconstructed density field is calculated using

$$\delta_r = -\nabla_{\mathbf{q}} \cdot \Psi(\mathbf{q}), \quad (4)$$

where we have used the same symbol  $\Psi$  to denote the displacement field but note that it is now a function of the Lagrangian coordinate  $\mathbf{q}$ , and the divergence is with respect to  $\mathbf{q}$  too. To calculate  $\Psi(\mathbf{q})$  on a regular  $\mathbf{q}$ -grid, we use the Delaunay tessellation field estimator code (DTFE; Schaap & van de Weygaert 2000; van de Weygaert & Schaap 2009; Cautun & van de Weygaert 2011), which is used to interpolate  $\Psi(\mathbf{x})$  to a regular  $\mathbf{q}$ -grid.

## 2.2 Reconstruction in redshift space

In observations, what is measured is the redshift-space coordinate,  $\mathbf{s}$ , of a particle (such as a galaxy), rather than the real-space position,  $\mathbf{x}$ . The two are related by

$$\mathbf{s} = \mathbf{x} + \frac{v_{\text{los}}}{aH(a)} \mathbf{n}, \quad (5)$$

where  $a$  is the scale factor,  $H(a)$  is the Hubble expansion rate at  $a$ ,  $\mathbf{n}$  is the line-of-sight (LOS) direction, and  $v_{\text{los}} = \mathbf{v} \cdot \mathbf{n}$  is the peculiar velocity of the galaxy along the LOS direction. As a result, galaxies infalling toward massive clusters or receding from void regions can cause redshift-space distortions – the RSD – to the isotropic spatial distribution they would have otherwise. For it to be practically useful, therefore, the reconstruction method described above must be extended to account for the RSD effect.

We remark that equation (1) contains only  $\mathbf{x}$  and  $\mathbf{q}$ . A similar equation that contains  $\mathbf{s}$  and  $\mathbf{q}$  may be obtained, allowing one to directly map between the  $\mathbf{s}$  and  $\mathbf{q}$  coordinates without having to worry about the  $\mathbf{x}$  coordinate. In other words, an equation similar to equation (3) can be written down, and the process depicted in Section 2.1 repeated, but with  $\mathbf{x}$  replaced by the observed coordinate



$s$ : with the assumption of no shell crossing, a unique solution of the  $s$ -to- $q$  mapping is still guaranteed. However, the derivative in the left-hand side of equation (3) is *formally* isotropic, whereas  $\delta(s)$  is anisotropic due to RSD, which means that the solutions  $\Theta(s)$  and  $\Psi(s)$  must be anisotropic. It is not clear whether this anisotropy would simply go away (as one would hope for) in the reconstructed density calculated using equation (4).

A different way to view this point is the following: our non-linear reconstruction method starts from a set of inhomogeneously distributed particles, and gradually moves the particles to a uniform distribution; in this process, particles can be moved in all directions as the algorithm sees necessary. If we knew how to correct  $s$  to get  $x$  exactly, the reconstruction would take two steps – first doing that correction to get  $x$  and then solving equations (3) and (4); in the first step particles are moved along the LOS direction only, whilst in the second step they are moved in all directions. If we attempt to directly map  $s$  to  $q$ , the first step in the above is omitted, and it is highly probable that the final solution obtained in this ‘crude’ way differs from that of the previous, ‘correct’, approach. One possible way to overcome this issue is to account for the additional displacements of galaxy LOS positions due to RSD by including extra terms in the equation for  $\Psi$  (Nusser & Davis 1994); the resulting equation at linear order can be solved in configuration space by finite difference (e.g. Padmanabhan et al. 2012) or in Fourier space by fast Fourier transform (FFT; see e.g. Burden et al. 2014). In the latter case, the extra term breaks the translational invariance of the problem, which prevents the use of simple FFT and leads to the development of schemes to improve the solution iteratively (e.g. Burden et al. 2015). As stated, these schemes are based on the solution to a linearized equation for  $\Psi$ , whilst we want to find a solution to the non-linear reconstruction equation derived from equation (3), and it is unclear how straightforward it is to generalize them here.

An alternative method is to keep using the  $x$  coordinate in the reconstruction equation (3), but add a conversion from  $s$  to  $x$  somewhere before that equation is solved. In the Zel’dovich approximation (ZA), the displacement field  $\Psi$  and peculiar velocity field  $v$  are related as

$$\frac{v(x)}{aH} = f \nabla \Phi_v = f \Psi[q(x)], \quad (6)$$

where  $\Phi_v$  is the velocity potential,  $f \equiv d \ln D_+ / d \ln a$  is the linear growth rate, and  $D_+$  the linear growth factor. This suggests that, in equation (5),  $s$  can be written as a function of  $x$  and  $\Theta(x)$  (the latter is the potential for  $\Psi$ ). However, the function that connects the three quantities –  $\delta(s)$ ,  $\delta(x)$ , and  $\Theta(x)$  – does not have an a priori known form, making it impossible to replace  $\delta(s)$  with  $\delta(x)$  in equation (3). This motivates a new iterative method here, which can be schematically summarized as

$$x^{(k+1)} = s - \frac{v^{(k)} \cdot n}{aH(a)} n \leftarrow v^{(k)} \leftarrow \Psi^{(k)} \leftarrow \delta^{(k)} \leftarrow x^{(k)}, \quad (7)$$

where  $k = 0, 1, 2, 3, \dots$  is the iteration number, and  $v^{(k)}$  the velocity field after the  $k$ th iteration, which is given by equation (6) with  $\Psi$  replaced by

$$\Psi^{(k)} \equiv \Psi(x^{(k)}), \quad (8)$$

i.e.  $\Psi^{(k)}$  is obtained by solving equation (3) using the particle coordinate after the  $k$ th iteration,  $x^{(k)}$ , to calculate the density field on the right-hand side:

$$\delta^{(k)} \equiv \delta(x^{(k)}). \quad (9)$$

At the first iteration step,  $k = 0$ , we simply set  $x^{(0)} = s$  as our ‘initial guess’, such that  $v^{(0)} = 0$ : this is equivalent to doing the reconstruction by assuming that the particles’ redshift-space coordinates are identical to their real-space coordinates. Note that in equation (7)  $s$  is the *observed* coordinate in redshift space, which is *fixed* during the iterations.

The first equality of equation (7) merits further comment as its simple form could obscure a subtle point, namely  $\Psi^{(k)}$ , and hence  $v^{(k)}$ , are evaluated at the  $x$ , which is not known prior to the reconstruction, rather than  $s$  coordinate. This indicates that in principle this is a non-linear equation for  $x$ . Accurate solution can be found numerically once the  $\Psi^{(k)}(x)$  or  $v^{(k)}(x)$  fields are known. This can be done as follows: start with the approximate solution from  $v^{(k)} \approx v^{(k)}(s)$  as an initial trial solution to  $x$ ,  $x_{j=1}$ , then take  $v^{(k)} \approx v^{(k)}(x_{j=1})$  to obtain an improved solution,  $x_{j=2}$ , and so on, until  $x_j$  converges for the galaxy considered under the velocity field  $v^{(k)}$ . Schematically, this can be viewed as an iterative procedure to solve the equation,

$$x_g^{(k+1)} = s_g - \frac{1}{aH} v^{(k)}(x_g^{(k)}) \cdot n, \quad (10)$$

where the explicit dependence of  $v^{(k)}$  on  $x^{(k+1)}$  makes it a non-linear algebraic equation for  $x^{(k+1)}$ , and we have used the subscript ‘g’ as a reminder that the coordinates are for galaxies. In practice, we used a simplified version of this scheme, described by

$$x_g^{(k+1)} = s_g - \frac{1}{aH} v^{(k)}(x_g^{(k)}) \cdot n, \quad (11)$$

namely the iterations for  $k$  and  $j$  are approximately done together. If convergence is achieved for  $x_g^{(k)}$ , we expect these two approaches to give consistent results with the latter one easier to implement. We plan to implement the full (iterative) solution to equation (10) as a future extension of the current pipeline.

## 2.3 Simulation

As a proof-of-concept study, in this paper we consider galaxy catalogues whose number density,  $n_g$ , and redshift match that of the BOSS CMASS data. More explicitly, the mock galaxy catalogues, first used in Cautun et al. (2018), were constructed using the halo occupation distribution (HOD) model and parameters given in Manera et al. (2013), and halo catalogues from N-body simulations of the  $\Lambda$ CDM model. The simulations were run using the RAMSES code (Teyssier 2002), employing  $1024^3$  particles in a cubic box of co-moving size  $1024 h^{-1}$  Mpc, and the cosmological parameters are

$$\{\Omega_m, \Omega_\Lambda, h, n_s, \sigma_8\} = \{0.281, 0.719, 0.697, 0.971, 0.8\}, \quad (12)$$

in which  $\Omega_m$  and  $\Omega_\Lambda$  are, respectively, the density parameters for matter and the cosmological constant ( $\Lambda$ ),  $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$  with  $H_0$  the Hubble constant,  $n_s$  is the primordial power spectrum index and  $\sigma_8$  denotes the rms matter density fluctuation smoothed on scales of  $8 h^{-1}$  Mpc. Further details of the simulations and of the HOD parameters are not very relevant for this paper, and so we opt to not report them here, but simply note that the galaxy catalogues correspond to redshift,  $z = 0.5$ , have a galaxy number density of  $n_g \simeq 3.2 \times 10^{-4} [h^{-1} \text{ Mpc}]^{-3}$ , and that RSD effects on the coordinates of our mock galaxies were implemented by displacing the galaxies, according to their peculiar velocities from the HOD, along the three axes of the simulation box, by adopting the distant observer approximation:<sup>1</sup> this means that for a given simulation we

<sup>1</sup>However, the method described here is not limited to the distant observer approximation.

have produced three HOD galaxy catalogues in redshift space. We have five independent realizations of simulations and therefore 15 galaxy catalogues at  $z = 0.5$ ; in the analysis of Section 3, we will only use the first galaxy catalogue, whilst all 15 are used in Section 4.

## 2.4 Implementation of the algorithm

The description of our iterative reconstruction algorithm in the previous subsection is quite schematic, and therefore in this subsection we give more technical details of its implementation. The presentation here shall follow the logic as depicted in the flowchart, Fig. 1, and for clarity we also list all the physical or numerical parameters, and their meanings, in Table 1.

The main ingredients of the reconstruction algorithm are listed below (where a superscript  $(k)$  denotes the corresponding quantities after the  $k$ th reconstruction iteration):

(i) *Creating the galaxy density field  $\delta_g^{(k)}$  on a uniform grid using the approximate real-space coordinates of the galaxies,  $\mathbf{x}^{(k)}$* : This is done using the triangular-shaped cloud (TSC) mass assignment scheme implemented in the DTFE public code (Cautun & van de Weygaert 2011). Note that we do *not* use actual Delaunay tessellation to calculate the density field, as it has been shown by Birkin et al. (2019) – and checked again in this project – that this leads to a poorer reconstruction performance.

The size of the uniform grid on which  $\delta_g^{(k)}$  is calculated has some effect on the reconstruction result, and in this work we have adopted a grid with  $512^3$  cells, i.e. with cell size  $\mathrm{d}x = 2h^{-1}$  Mpc, because using a grid with even higher resolution does not make a significant difference (Birkin et al. 2019).

(ii) *Calculating the displacement field  $\Psi$  and performing reconstruction*: Here, things become a bit tricky: even though we are trying to simultaneously do reconstructions of the initial density field and the real-space galaxy coordinates, the optimal technical specifications are not the same in the two cases. As a result, we actually do two reconstruction calculations of  $\Psi$  for a given  $\delta_g^{(k)}$  field, both using the ECOSMOG code developed by Shi et al. (2018) and Birkin et al. (2019).

In the first calculation, the objective is to undo the RSD and thus to bring the galaxy coordinates,  $\mathbf{x}^{(k)}$ , closer to their true real-space values,  $\mathbf{x}$ . Here, our concern is that the stretching effects of FoG could lead to erroneous estimation of the large-scale density field, causing worse performance of the method. To reduce its impact, we follow Hada & Eisenstein (2018) and calculate the density field,  $\delta_g^{(k)}$ , using an anisotropic smoothing function. The filtering function is chosen to be a skewed Gaussian that has a different smoothing length along the line-of-sight direction, and the smoothed galaxy density field is given, in Fourier space, as<sup>2</sup>

$$\tilde{\delta}_{g,S}^{(k)}(k) = \tilde{\delta}_g^{(k)} \tilde{G}(\mathbf{k}) \equiv \tilde{\delta}_g^{(k)} \exp \left[ - (k_n^2 S_n^2 + k_p^2 S_p^2) \right], \quad (13)$$

where  $\mathbf{k}$  is the wavenumber with  $k_n$  and  $k_p$  representing the wave numbers along the line of sight and perpendicular to it. The functions  $\tilde{\delta}_g^{(k)}$ ,  $\tilde{G}(\mathbf{k})$  are the Fourier transformations of  $\delta_g^{(k)}$  and the filter mentioned above. This introduces two extra parameters for the algorithm,  $S_n$  and  $S_p$ , and in what follows we express them by  $S = S_p$  (the smoothing length perpendicular to LOS) and a dimensionless parameter  $C_{\text{ani}} \equiv S_n/S$ , with  $C_{\text{ani}} > 1$  representing a larger smoothing length along the LOS. The calculation from here on is similar as

before, but with  $\delta_{g,S}^{(k)}$  instead of  $\delta_g^{(k)}$  being fed into ECOSMOG, and  $b^{(k)}$  is applied again to convert this to an approximated non-linear matter density field.<sup>3</sup> The displacement field obtained here is denoted as  $\Psi_S^{(k)}$ , from which we can derive the ‘improved’ real-space galaxy coordinates,  $\mathbf{x}^{(k+1)}$ , as

$$\mathbf{x}^{(k+1)} = \mathbf{s} - f \Psi_S^{(k)}, \quad (14)$$

where  $f$  is the linear growth rate introduced above, which we take as a scale-independent (but time-dependent) constant, as is the case for  $\Lambda$ CDM and several dark energy and modified gravity models. At  $z = 0.5$ , the equation

$$f(z) \simeq [\Omega_m(z)]^{0.55} \quad (15)$$

is a very good approximation, which gives a value of  $f = 0.735$ , in good agreement with numerical result obtained by using the cosmological parameters given above. However, in the actual calculation we have left  $f$  to be a free parameter to be varied because its value is a priori unknown in observations.

In the second calculation, the aim is to obtain the reconstructed matter density field,  $\delta_{\text{rec}}^{(k)}$ , using the relation

$$\delta_{\text{rec}}^{(k)} = -\nabla \cdot \Psi^{(k)}, \quad (16)$$

where the displacement field at the  $k$ th iteration,  $\Psi^{(k)}$ , is calculated by applying ECOSMOG to  $\delta_g^{(k)}/b^{(k)}$ , without doing any smoothing (which would degrade the performance; see below and Birkin et al. 2019). Here,  $b^{(k)}$  is the linear bias parameter such that  $\delta_g^{(k)}/b^{(k)}$  is an approximation to the non-linear matter density field; note that here we assume different values of  $b^{(k)}$  need to be used in the different iterations.

(iii) *Checking for convergence*. As an iterative solution scheme, we need a criterion (or a set of criteria) to decide when the iterations can be stopped. Usually, convergence is deemed to be achieved if the error (defined in whatever way) is reduced to below some preset tolerance, e.g. some small number. The problem at hand is more complicated in that, a priori, there is no ‘target’ solution to be used to clearly define the ‘error’. Therefore, here we opt for a set of loose criteria for convergence:

C1: A set of estimators obtained from the reconstruction outcome ‘stabilize’ and do not change further with increasing number of iterations ( $k$ ). This is a generic convergence criterion which is essential for the method to work, and we require it to be satisfied for any estimator to be considered. This criterion is also practically useful, as it applies to both statistics extracted directly from observations (such as estimator E3 to be introduced below) and theoretical quantities that are only known in controlled experiments, such as simulations. The latter, however, are also helpful since they offer other ways to assess the performance of and to determine the optimal parameters for the reconstruction; for this reason, we also introduce two more convergence criteria that apply only to theoretical quantities.

C2: Assuming that convergence is achieved after iteration  $k = K$ , then reconstructed matter density field  $\delta_{\text{rec}}^{(K)}$  is ‘closer’ to the initial density field  $\delta_{\text{ini}}$  than any of the pre-convergence results,  $\delta_{\text{rec}}^{(k)}$ ,  $\forall k < K$ ; here,  $\delta_{\text{ini}}$  is a theoretical quantity.

C3: The reconstructed galaxy coordinates  $\mathbf{x}^{(K)}$  are ‘closer’ to the true real-space galaxy coordinates  $\mathbf{x}$  than any pre-convergence results,  $\mathbf{x}^{(k)}$ ,  $\forall k < K$ ; here,  $\mathbf{x}$  is a theoretical quantity.

<sup>2</sup>Note the slight abuse of notation here:  $k$  is used both to denote the iteration number and to represent the wavenumber/vector in Fourier space.

<sup>3</sup>In principle, the  $b^{(k)}$  parameters used here can be different from the ones used in the first calculation above, but in our implementation, we have used the same  $b^{(k)}$  for a given  $k$  iteration.

It is not our objective to be very quantitative in defining convergence, and instead we simply check that ‘by eye’, i.e. we stop the iterations if the statistic or estimator of interest has stabilized and does not change significantly after further iterations. Four estimators are defined, which can be constructed from the reconstruction outcome, to allow us to test these criteria. Different estimators may need different numbers of iterations before convergence, and these are shown in Table 1.

For Criterion C2, we use the usual cross-correlation coefficient,  $r$ , between the reconstructed and initial density fields, to characterize the similarity between them. The correlation coefficient between any two fields  $\delta_a$  and  $\delta_b$  is defined as

$$r[a, b] \equiv \frac{\tilde{\delta}_a \tilde{\delta}_b^* + \tilde{\delta}_a^* \tilde{\delta}_b}{2\sqrt{\tilde{\delta}_a \tilde{\delta}_a^*} \sqrt{\tilde{\delta}_b \tilde{\delta}_b^*}}, \quad (17)$$

where  $\tilde{\delta}_a$  and  $\tilde{\delta}_b$  are the Fourier transforms of  $\delta_a$  and  $\delta_b$  and a superscript \* denotes taking the complex conjugate. A value of  $r[a, b] = 1$  means perfect correlation, whilst  $r[a, b] = 0$  means that  $a$  and  $b$  are completely random. In other words, for C2 we would like that  $r[\delta_{\text{rec}}^{(K)}, \delta_{\text{ini}}]$  to be closer to 1 than  $r[\delta_{\text{rec}}^{(k)}, \delta_{\text{ini}}]$ , for  $\forall k < K$ . Since  $r[a, b]$  is a function of scale, or Fourier wave number,  $k$ , ideally we hope the above applies for all wavenumber values or, if that is not possible, at least for the range of wavenumbers of most interest to us.

For C3 we have defined a similar estimator by cross-correlating  $\delta_g^{(k)}$  with the final real-space galaxy density field,  $\delta_g^r$ , and requiring that  $r[\delta_g^{(K)}, \delta_g^r]$  is closer to 1 than  $r[\delta_g^{(k)}, \delta_g^r]$ ,  $\forall k < K$ .

We have also defined two more estimators based on the argument that, if  $\mathbf{x}^{(K)}$  is close enough to  $\mathbf{x}$ , then the two-point correlation functions obtained from these two galaxy catalogues should also be close to each other. In particular, the RSD-induced anisotropy in the two-point correlation function of the redshift-space ( $s$ ) galaxy catalogue should be largely removed in the reconstructed,  $\mathbf{x}^{(K)}$ , galaxy catalogue. Therefore, we require that  $\xi_2[\mathbf{x}^{(K)}]$ , the quadrupole of the two-point galaxy correlation function of the  $\mathbf{x}^{(K)}$  catalogue, be closer to 0 than  $\xi_2[\mathbf{x}^{(k)}]$ ,  $\forall k < K$ .<sup>4</sup>

In addition, we would also expect that  $\xi_0[\mathbf{x}^{(K)}]$ , the monopole of the two-point galaxy correlation function of the  $\mathbf{x}^{(K)}$  catalogue, to be close to the real-space galaxy correlation function  $\xi_{\text{gg}}$ . Therefore, a further requirement is that the ratio  $\xi_0[\mathbf{x}^{(K)}]/\xi_{\text{gg}}$  be closer to 1 than  $\xi_0[\mathbf{x}^{(k)}]/\xi_{\text{gg}}$ ,  $\forall k < K$ . In this paper, we measure  $\xi_0$  and  $\xi_2$  using the publicly available code ‘Correlation Utilities and Two-point Estimators’ (CUTE; Alonso 2012).

Note that in certain situations we may need to loosen the above requirements. Taking  $\xi_2[\mathbf{x}^{(k)}]$  for example, it is possible that for some intermediate  $k < K$  the result coincidentally gets very close to zero (this may happen if  $\xi_2[\mathbf{x}^{(k)}]$  oscillates around 0 for increasing  $k$ ). Therefore, it is always safe to try a couple more iterations even if the result seems to have converged.

(iv) *Finalizing the code.* Finally, once convergence is deemed to have been achieved, we stop the iterations at  $k = K$ .

In what follows, to avoid carrying cumbersome notations everywhere, we shall call the four estimators introduced above E1, E2, E3, and E4, respectively. Note that out of these estimators, only E3

is applicable in real observations because the other three all require something that only exists in simulations in their definitions –  $\delta_{\text{ini}}$  for E1,  $\delta_g^r$  for E2, and  $\xi_{\text{gg}}(r)$  for E4. As a result, the latter estimators are mainly used in this work as theoretical tools to demonstrate the performance of the iterative reconstruction algorithm and to determine the optimal technical parameters.

On the other hand, E3 can be estimated using observational data alone. Therefore, our objective in the following parts of this paper is to check what is the potential of using E3 alone to determine the ‘best-fitting’ values of the physical parameters, such as  $f$  and  $b^{(k)}$ , and to do the RSD reconstruction. If  $f, b$  could be precisely determined in this process, then that would be an additional benefit of this new algorithm, along with simultaneously giving us approximate reconstructions of the initial (linear) and final (non-linear) matter density fields and the final real-space galaxy density field (or coordinates). These will turn out to be very useful information as we exemplify and discuss later. In the less ideal scenario, if  $f, b$  could not be accurately determined (for example because the reconstruction outcome is not very sensitive to them), then the other benefits would remain.

Note that we can also use higher-order multipole moments, such as the hexadecapole  $\xi_4[\mathbf{x}^{(k)}](s)$ , as more estimators to check the convergence, namely  $\xi_4[\mathbf{x}^{(K)}]$  must be closer to 0 than  $\xi_4[\mathbf{x}^{(k)}]$ ,  $\forall k < K$ . These have the advantage that they can be obtained from real observational data. In particular, it would be interesting to see if they offer consistent (or complementary) constraints on the physical parameters, such as  $f$  and  $b$ . However, for our galaxy catalogues the number density  $n_g \simeq 3.2 \times 10^{-4} [h^{-1} \text{Mpc}]^{-3}$  is too low and the measurements of  $\xi_4[\mathbf{x}^{(k)}]$  are too noisy. Therefore, we shall leave a check of the impacts of such additional estimators to a future work, where we’ll test the reconstruction algorithm using galaxy catalogues with various number density cuts.

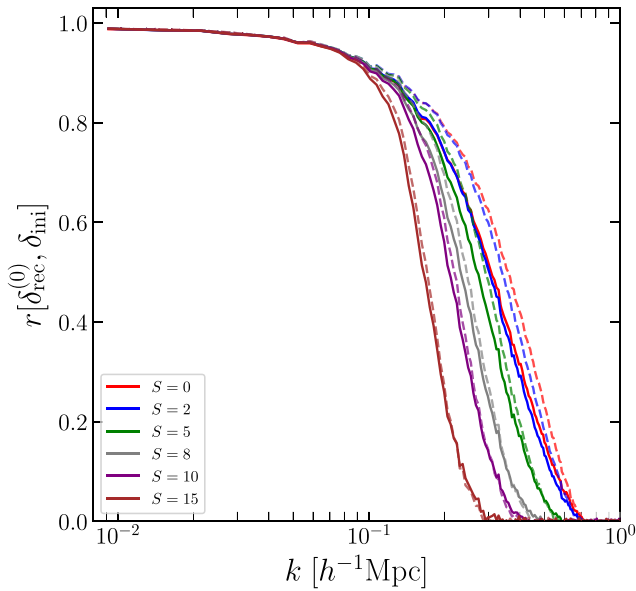
### 3 RECONSTRUCTION TESTS AND PERFORMANCE

We tested the reconstruction pipeline for a large number of combinations of the physical and technical parameters,  $\{f, b^{(k)}, S, C_{\text{ani}}\}$ , in which  $b^{(k)}$  were allowed to vary with the iteration number,  $k$ , in order to settle to the most optimal choices of  $S, C_{\text{ani}}$  and to explore the potential of constraining  $f, b$  as a byproduct of reconstruction. The optimal values for these parameters are summarized in the last column of Table 1, and in this section, we will show the impacts of varying these parameters on the reconstruction performance. As we have a relatively large parameter space, we shall only vary a subset of them – whilst fixing the others to the optimal values – at a given time.

Before going to the details, in Fig. 2 we present a quick visual inspection of the impact of RSD on the reconstruction performance. The red dashed line is estimator E1,  $r[\delta_{\text{rec}}, \delta_{\text{ini}}]$ , between the initial matter density field,  $\delta_{\text{ini}}$ , and the reconstructed matter density field,  $\delta_{\text{rec}}$ , from the final galaxy catalogue in real space. The red solid line differs by replacing  $\delta_{\text{rec}}$  with  $\delta_{\text{rec}}^{(0)}$ , which is the reconstructed matter density field from the zeroth-iteration of our RSD reconstruction, namely by incorrectly assuming that the redshift-space coordinates of the galaxies are also their real-space coordinates without any corrections, or equivalently applying the reconstruction code of Birkin et al. (2019) directly to our redshift-space galaxy catalogue without using iterations. We can see that not cleaning up RSD effects causes the correlation to become smaller than in real-space reconstruction, which is as expected. However, the impact is mild, which is perhaps because of the relatively low galaxy number density

<sup>4</sup>Note that we have used  $[\ ]$  to highlight that  $\mathbf{x}^{(k)}$  is not an argument of  $\xi_2$  but simply is a symbol to represent a given galaxy catalogue. The proper argument for  $\xi_2(s)$ , not shown here to lighten the notation, is the galaxy pair distance in redshift space,  $s$ . As above, ideally we would like  $\xi_2[\mathbf{x}^{(K)}]$  to be close to 0 on all scales or, if it is not possible, at least in the scales of most interest to us.





**Figure 2.** The cross-correlation coefficients of the initial density field with the reconstructed matter density field from the real-space galaxy catalogue ( $r[\delta_{\text{rec}}, \delta_{\text{ini}}]$ ; dashed lines) and with the reconstructed matter density field from the redshift-space galaxy catalogue ( $r[\delta_{\text{rec}}^{(0)}, \delta_{\text{ini}}^{(0)}]$ ; solid lines) using *no* iterations. The various coloured lines correspond to the results for which the galaxy density field has been smoothed by a skewed Gaussian filter with  $C_{\text{ani}} = 1.0$  and  $S = 0$  (no smoothing; red; rightmost curve), 2, 5, 10, and  $15 h^{-1} \text{ Mpc}$  (brown; leftmost curve). The bias parameter used here is  $b^{(0)} = 2.0$ .

used here. As a result, we expect that any improvement by iterative reconstruction will be mild as well (but note that both conclusions might not hold for galaxy catalogues with much higher  $n_g$ ).

The dashed and solid lines with other colours in Fig. 2 are very similar, but they correspond to results where both the real- and the redshift-space galaxy density fields are further smoothed – after the TSC mass assignment – using the skewed Gaussian filter described above, with  $C_{\text{ani}} = 1$ ,  $S = 2$  (blue), 5 (green), 8 (grey), 10 (purple), and  $15 h^{-1} \text{ Mpc}$  (brown). Notice that the red lines described above are results from unsmoothed galaxy density field and correspond to  $S = 0$ . We can see a clear trend that smoothing the galaxy density field leads to poorer outcomes of the reconstruction (as mentioned earlier), which is because the smoothing effectively suppresses the small-scale features of the density field. This is why when describing the flowchart (Fig. 1) above we emphasized that smoothing is used in calculating the displacement field  $\Psi_S^{(k)}$  which is needed to correct galaxy coordinates, and not in calculating the displacement field  $\Psi^{(k)}$  which is used to obtain the reconstructed matter density field. Also note that for all tests in Fig. 2 we have used  $b^{(0)} = 2.0$  and that  $f$  is not used here.

Another interesting feature in Fig. 2 is that, as the smoothing length  $S$  increases, the difference between real and redshift-space reconstructions reduces, and with  $S = 15 h^{-1} \text{ Mpc}$  (brown lines) the two cases almost agree perfectly with each other. This is again not surprising given that the effect of RSD is to shift galaxy positions whilst smoothing to certain extent undoes that shift. However, this is at a price of suppressing small-scale features and leading to poorer reconstruction results for both real and redshift spaces.

We now present the results of a wide range of tests to illustrate the (lack of) impacts of varying different physical and technical parameters used in the iterative procedure on the estimators defined

in the previous section. As mentioned above, these parameters serve both as fitting parameters used to identify the optimal reconstruction specifications, as well as informative vehicles that can provide valuable insights into the formation of large-scale structures.

### 3.1 Smoothing parameters $S$ and $C_{\text{ani}}$

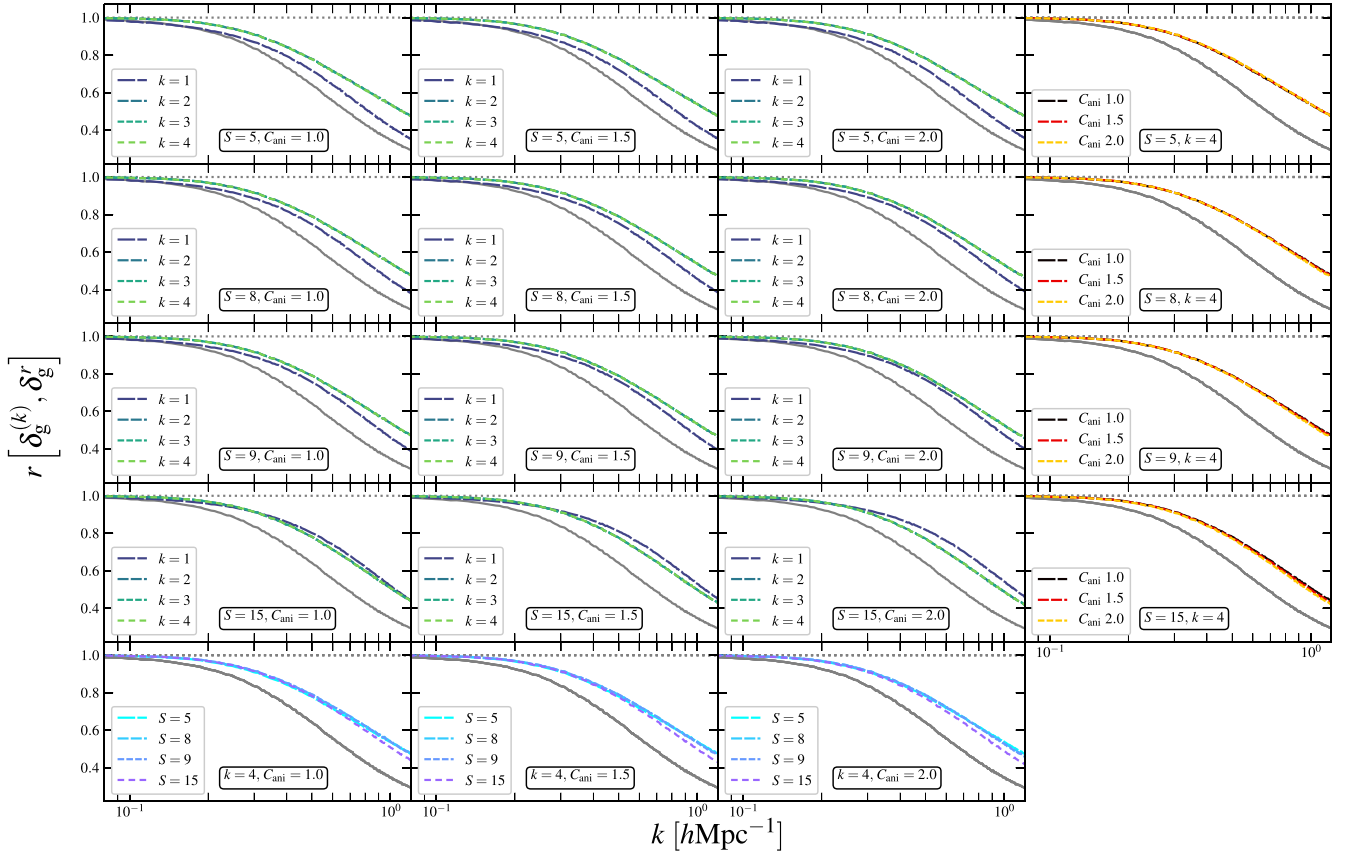
Let us first discuss the smoothing-related parameters  $S$  and  $C_{\text{ani}}$ .  $S$  represents the overall smoothing length, whilst  $C_{\text{ani}}$  characterizes the amount of anisotropic smoothing, with  $C_{\text{ani}} = 1.0$  indicating no anisotropy in the smoothing function and  $C_{\text{ani}} > 1.0$  indicating a longer smoothing length along the line-of-sight direction to suppress the impact of the FoG effect.

In Fig. 3, we compare estimator E2 constructed from the reconstruction outputs for  $4 \times 3$  combinations of  $(S, C_{\text{ani}})$ : four choices of  $S = 5, 8, 9$ , and  $15 h^{-1} \text{ Mpc}$  – and three choices of  $C_{\text{ani}} = 1.0, 1.5$ , and  $2.0$ . The results for each of the 12 combinations are shown in one of the 12 subpanels on the top left of Fig. 3, and in each subpanel, the different lines are the results after different numbers of iterations,  $k$ ,  $r[\delta_g^{(k)}, \delta_g^{(k)}]$ , with  $k = 1, 2, 3, 4$ . As a comparison, the dashed line represents  $r[\delta_g^s, \delta_g^r]$ , namely the cross-correlation between the final real and redshift-space galaxy density fields. Within each row, the smoothing scale perpendicular to the LOS,  $S_p$ , is fixed whilst  $S_n$ , as being the product of  $S$  and  $C_{\text{ani}}$ , changes across the columns.

By comparing the different columns in a given row in Fig. 3, it is evident that the effect of  $C_{\text{ani}}$  on estimator E2 is only significant for first few iterations. For  $k = 4$ , the difference between  $C_{\text{ani}} = 1.0, 1.5$ , and  $2.0$  is much smaller. The convergence criterion C1 is satisfied by all tests, regardless of the value of  $C_{\text{ani}}$ .

The overall behaviour of  $r[\delta_g^{(k)}, \delta_g^{(k)}]$  for small smoothing scale is as expected. One can consider the FoG effect as some ‘redistribution’ of galaxies around the centres of their host haloes, where virial motions of the former can lead to the measured galaxy coordinates differing from their actual values by an amount much larger the radii of the dark matter haloes. If uncorrected, this could cause a galaxy 1 which is closer to us than another galaxy 2 in real space to actually appear to be farther away than galaxy 2 in redshift space. In other words, a ‘shell crossing’ takes place due purely to the use of redshift space coordinates, and this violates one of the basic assumptions of the reconstruction method, which leads to a degraded performance of the latter. This impact can be alleviated if the galaxy density field is smoothed using a large filter, whose size is at least comparable to the typical peculiar-velocity-induced changes of galaxy distances in redshift space. Choosing a too large smoothing scale reduces the correlation due to the loss of information, so an optimal scale must be found that balances the FoG effect without compromising the accuracy of the reconstruction.

The physical reasoning given in the above paragraph is supported by the following observation of Fig. 3, namely in the cases of smaller smoothing lengths, for both  $S$  and  $C_{\text{ani}}$ , the cross-correlation  $r[\delta_g^{(k)}, \delta_g^{(k)}]$  is generally larger. Since the smoothing length  $S_n$  along the LOS direction is the product of  $S$  and  $C_{\text{ani}}$ , we have  $S_n = 9$  for  $(S, C_{\text{ani}}) = (9, 1)$  which are sufficiently large to smooth out the FoG effect (notice that for typical galaxies the LOS velocities are smaller than  $2000 \text{ km s}^{-1}$  so that  $\mathbf{v} \cdot \mathbf{n}/aH \lesssim 20 h^{-1} \text{ Mpc}$ ). Overall, we find that for a fixed  $C_{\text{ani}}$  value, increasing  $S$  (or equivalently  $S_p$ ) leads to a slightly faster convergence for E2 (as can be observed when comparing the first and fourth rows in Fig. 3); however, the final reconstruction is slightly worse (see the last row in Fig. 3); the same happens for increasing  $C_{\text{ani}}$  at fixed  $S$ , but the effect is much smaller, as indicated by the rightmost column of this figure; this is likely because our mock galaxy catalogues have a small fraction of



**Figure 3.** Estimator E2,  $r[\delta_g^{(k)}, \delta_g^r](k)$ , for various combinations of technical parameters  $S$  and  $C_{\text{ani}}$ . Each of the first four rows includes tests using a fixed  $S$ , which takes value of 5, 8, 9, and  $15 h^{-1}$  Mpc, respectively; each of the first three columns corresponds to tests using a fixed  $C_{\text{ani}}$ , which takes value of 1.0, 1.5, and 2.0, respectively. The  $4 \times 3$  block of subpanels on the top left show how E2 changes with increasing number of iterations for a given  $(S, C_{\text{ani}})$ . Each of the three subpanels at the bottom compares the results for fixed  $C_{\text{ani}}$  and varying  $S$ , at the last iteration; each of the four subpanels on the far right compares the results for fixed  $S$  and varying  $C_{\text{ani}}$ , again at the last iteration. The sparseness of the dashes lines increases with  $k$ ,  $S$ , or  $C_{\text{ani}}$  in the three different regions (see the legends). The grey solid lines are the same in all subpanels and show  $r[\delta_g^s, \delta_g^r](k)$ , which is the cross-correlation between the final galaxy density fields in redshift and real spaces. The grey dotted lines are zero.

satellites ( $\sim 11$  per cent), so that the FoG effect is not particularly important.

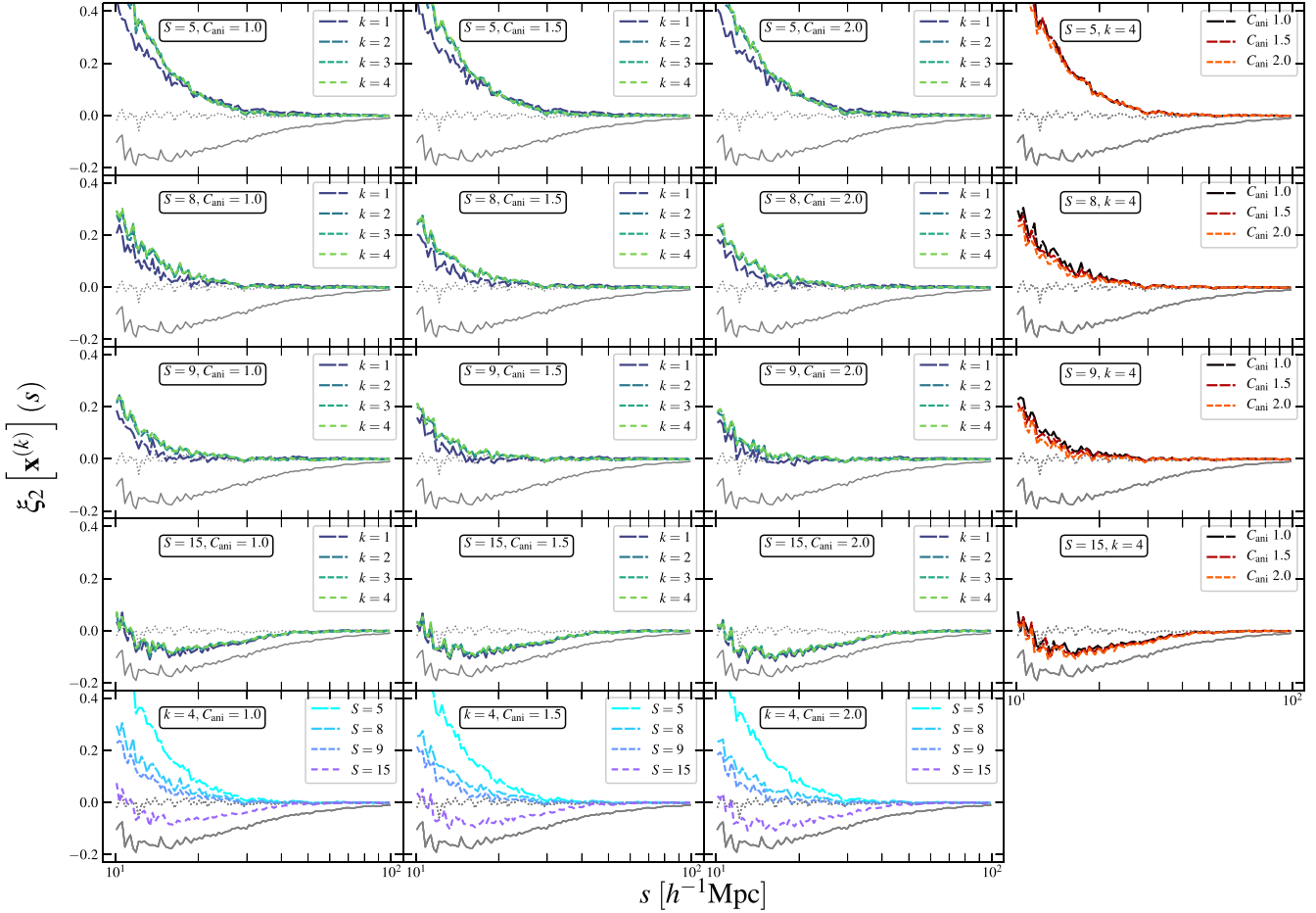
We next move on to estimator E3. Fig. 4 shows the quadrupole moments of reconstructed galaxy catalogues,  $\xi_2[\mathbf{x}^{(k)}](s)$ , for the same  $(S, C_{\text{ani}})$  parameter combinations as in Fig. 3. The two grey dashed lines, which are the same in all subpanels, are, respectively, the quadrupole moments measured from the final galaxy catalogues at  $z = 0.5$  in real (upper) and redshift (lower) space, and as expected the former is zero on all scales probed here ( $r \gtrsim 10 h^{-1}$  Mpc), whilst the latter is negative as a result of the Kaiser effect.

There are a few features in Fig. 4 that are noticeable. First of all, we find rapid and monotonic convergences, with the solutions generally never requiring more than just a few iterations for all the parameter combinations. The convergence becomes slower for small smoothing lengths ( $S = 5 h^{-1}$  Mpc), but the differences are minor. Secondly, unlike for E2, here the choice of  $S$  can have a significantly greater impact on the converged result of  $\xi_2[\mathbf{x}^{(K)}]$ : in the better scenarios, such as  $(S, C_{\text{ani}}) = (9 h^{-1} \text{ Mpc}, 1.0)$ , we can observe that  $\xi_2[\mathbf{x}^{(K)}](s) \simeq 0$  for  $s \gtrsim 20 h^{-1}$  Mpc, whilst in the less good cases, such as  $(S, C_{\text{ani}}) = (15 h^{-1} \text{ Mpc}, 2.0)$  this can only be achieved at  $s \gtrsim 50 h^{-1}$  Mpc. Thirdly, overall speaking, if the smoothing length  $S$  is too large, there is insufficient correction to make  $\xi_2[\mathbf{x}^{(K)}]$  go to zero on all but the largest scales ( $s \gtrsim 50 h^{-1}$  Mpc), whilst if  $S$

is too small, the correction seems to ‘overshoot’ and make  $\xi_2[\mathbf{x}^{(K)}]$  positive. This can be reasonably explained, given that oversmoothing (i.e. a too large  $S$ ) would lead to  $\Psi_S^{(k)}$  values which are appropriate only for large scales and therefore the resulting corrections to galaxy coordinates are not enough on small scales, whilst in the case of undersmoothing (a too small  $S$ ) the resulting values of  $\Psi_S^{(k)}$  can be strongly affected by structures on very small scales, causing ‘too much’ correction. Finally, for a specific  $S$ , varying  $C_{\text{ani}}$  between 1.0 and 2.0 does not seem to have a significant impact on the converged result of E3 (after four iterations, see the right column of Fig. 4).

Fig. 5 is similar to Fig. 4, but shows the impact of  $(S, C_{\text{ani}})$  on estimator E4, i.e.  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{\text{gg}}(r)$ . The convergence properties are comparable to the case of E3, with convergence achieved after two to four iterations in all cases, and the observation in the cases of E2 and E3 that  $C_{\text{ani}}$  has a negligible effect holds here as well. The result is again sensitive to  $S$ , with a value of  $S$  that is too small producing insufficient correction to bring E4 to 1.0 on all scales, whilst an  $S$  value that is too large causes an incorrect shape of E4 as a function of  $s$  by deviating it from a constant value in  $s$ . Overall, we find that  $S = 8-9 h^{-1}$  Mpc is capable of bringing E4 closest to 1.0 on all scales  $s \gtrsim 20 h^{-1}$  Mpc.

Very reassuringly, in general, for combinations  $(S, C_{\text{ani}})$  that bring  $\xi_2[\mathbf{x}^{(K)}]$  closer to zero down to small  $s$  values, the corresponding



**Figure 4.** The same as Fig. 3, but for estimator E3,  $\xi_2[x^{(k)}](s)$ . The grey dotted and solid lines are the same in all subpanels: the former, which is very close to 0 on the entire range of scales, is the quadrupole moment measured from the real-space galaxy catalogue at  $z = 0.5$ , whilst the latter, which is negative in the whole  $s$  range, is that measured from the redshift-space galaxy catalogue at  $z = 0.5$ .

$\xi_2[x^{(k)}]/\xi_{gg}(r)$  curves are also close to 1.0, which suggests that the reconstruction can get the two correct simultaneously (as it is expected to).

To summarize, we find that

- (i) estimator E2 prefers a larger smoothing length  $S$ , and the value of  $C_{\text{ani}}$  is not as important;
- (ii) compared with the E2 estimator, E3 and E4 are more sensitive to  $S$ , and disfavour either very large or very small values of  $S$ ;
- (iii) the key objective of the reconstruction algorithm, namely to accurately remove the RSD effects (or equivalently to bring E3 to 0 and E4 to 1), can be achieved for  $S \sim 8\text{--}9 h^{-1} \text{ Mpc}$  and  $C_{\text{ani}} = 1.0$ .

These have motivated us to choose  $9 h^{-1} \text{ Mpc}$  as the optimal value for  $S$  (for galaxy number density  $n_g = 3.2 \times 10^{-4} [h^{-1} \text{ Mpc}]^{-3}$ ). As for  $C_{\text{ani}}$ , given its weak impact on all estimators, we opt for the simple choice by setting its default value to 1.0. This simplifies our pipeline as it now adopts isotropic smoothing, but note that for other  $n_g$  values this needs to be checked separately.

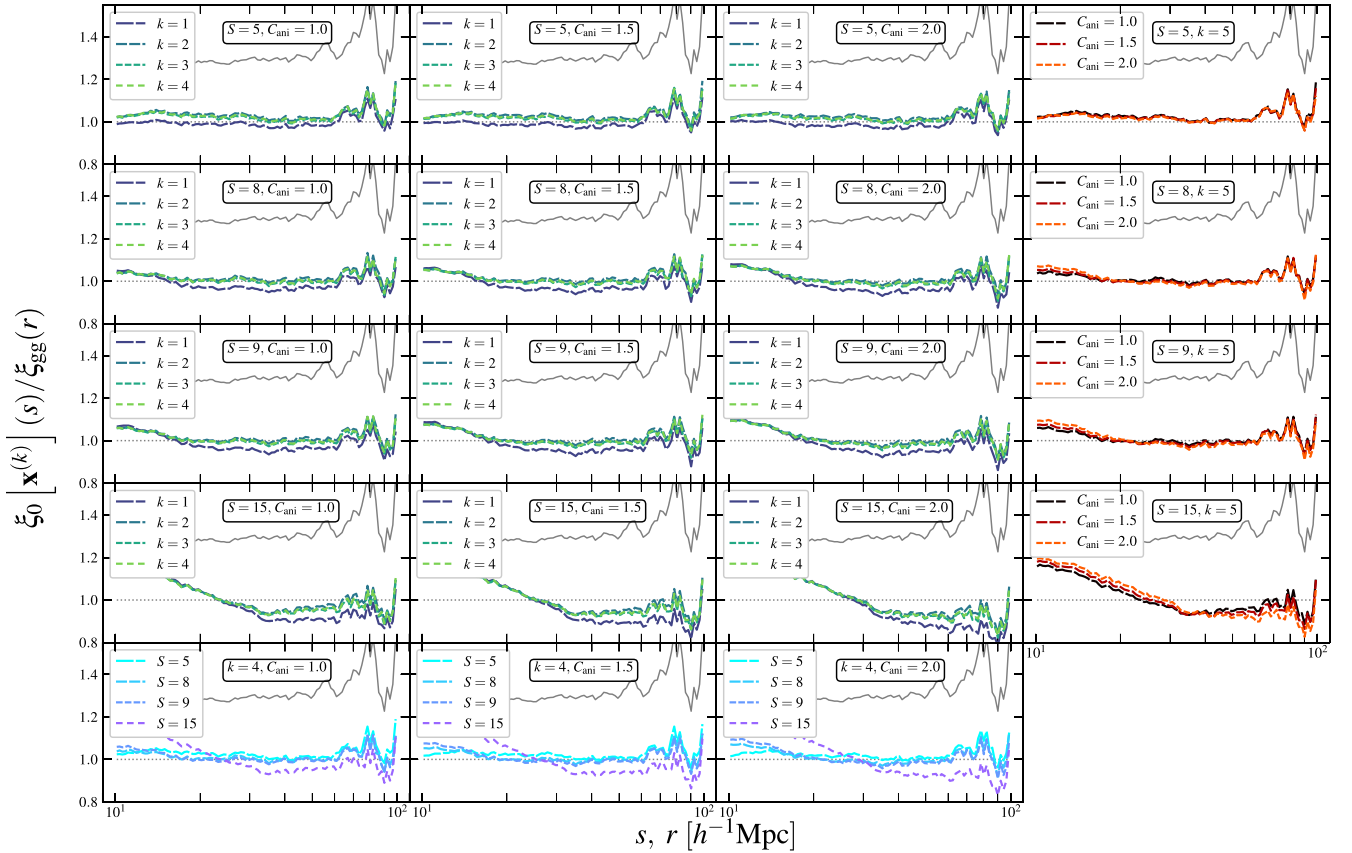
The results for E1 are very similar to E2, and to avoid getting this paper too heavy on technical details, we refrain from showing them here.

### 3.2 Galaxy bias parameter $b^{(k)}$

Next, we explore the impact on reconstruction of the linear galaxy bias parameter,  $b^{(k)}$ . As mentioned above, this parameter is used to convert a non-linear galaxy density field to a non-linear matter density field, since it is the latter that enters the reconstruction equation (Shi et al. 2018; Birkin et al. 2019). In the  $\Lambda$ CDM scenario, linear bias is time dependent but scale independent on large and linear scales. Therefore, given that we work at a fixed redshift,  $z = 0.5$ , we simply take  $b^{(k)}$  as a constant number.

Whilst the linear galaxy bias is a physical parameter, we do not necessarily know its value accurately as this depends on the galaxy population. This is especially true in observations, where we do not even have precise knowledge of the cosmological parameters. As a result, by trying different values of  $b^{(k)}$ , we can test whether the exact value adopted is important – if yes, then the reconstruction can be used to determine this value; if not, then not precisely knowing its value would not impact the reconstruction outcome strongly.

We do, on the other hand, allow the bias  $b^{(k)}$  to vary between the different iterations in the reconstruction process – and there is a reason for this. Usually, when speaking about galaxy bias, one refers to the bias in real space,  $\delta_g^r \equiv b\delta_m$ , where  $\delta_g^r$  and  $\delta_m$  are the density contrasts of galaxies and matter in real space, respectively. However, at a given iteration of reconstruction, especially when  $k = 0$ , what we



**Figure 5.** The same as Fig. 3, but for estimator E4,  $\xi_0[\mathbf{x}^{(k)}(s)]/\xi_{\text{gg}}(r)$ . The straight dotted and the wiggly solid grey lines are the same in all subpanels: the former is the constant 1.0 to guide the eyes, whilst the latter is the ratio between the monopole moments measured from the redshift- and real-space galaxy catalogues at  $z = 0.5$ .

have are not exactly the galaxy coordinates in real space but some approximations (for  $k > 0$ ), or their coordinates in redshift space (for  $k = 0$ ). Therefore, to convert the galaxy density field  $\delta_g^s$  or  $\delta_g^{(k)}$  to the real-space matter density field, an additional bias is needed and this additional bias depends on how much deviation  $\delta_g^s$  or  $\delta_g^{(k)}$  has with respect to the real-space galaxy density field,  $\delta_g^r$ . One can argue that  $b^{(0)}$  should be the largest because the additional bias will apply to  $\delta_g^s$  that differs most from  $\delta_g^r$ , whilst for  $k > 0$  the additional bias correction required should decrease as  $\delta_g^{(k)}$  gets closer to  $\delta_g^r$ .<sup>5</sup>

For this reason, in the tests here we sample a  $2 \times 3 \times 3$  grid of the parameter space, with  $b^{(0)} \in \{2.3, 2.0\}$ ,  $b^{(1)} \in \{1.8, 2.0, 2.2\}$ , and  $b^{(2)} \in \{1.9, 2.0, 2.1\}$ . For further iterations ( $k \geq 3$ ), we simply fix  $b^{(k)} = 2.0$  because, as we shall see shortly, whilst differences can be spotted between  $b$  being 1.8, 2.0, and 2.2, it is very mild between 1.9, 2.0, and 2.1. The other reconstruction and physical parameters are fixed to  $S = 9.0 h^{-1} \text{Mpc}$ ,  $C_{\text{ani}} = 1.0$ , and  $f = 0.735$  for the  $b^{(k)}$  tests presented in this subsection.

The test results for the E3 estimator (the quadrupole moment) are presented in Fig. 6, which demonstrate how marginal the differences

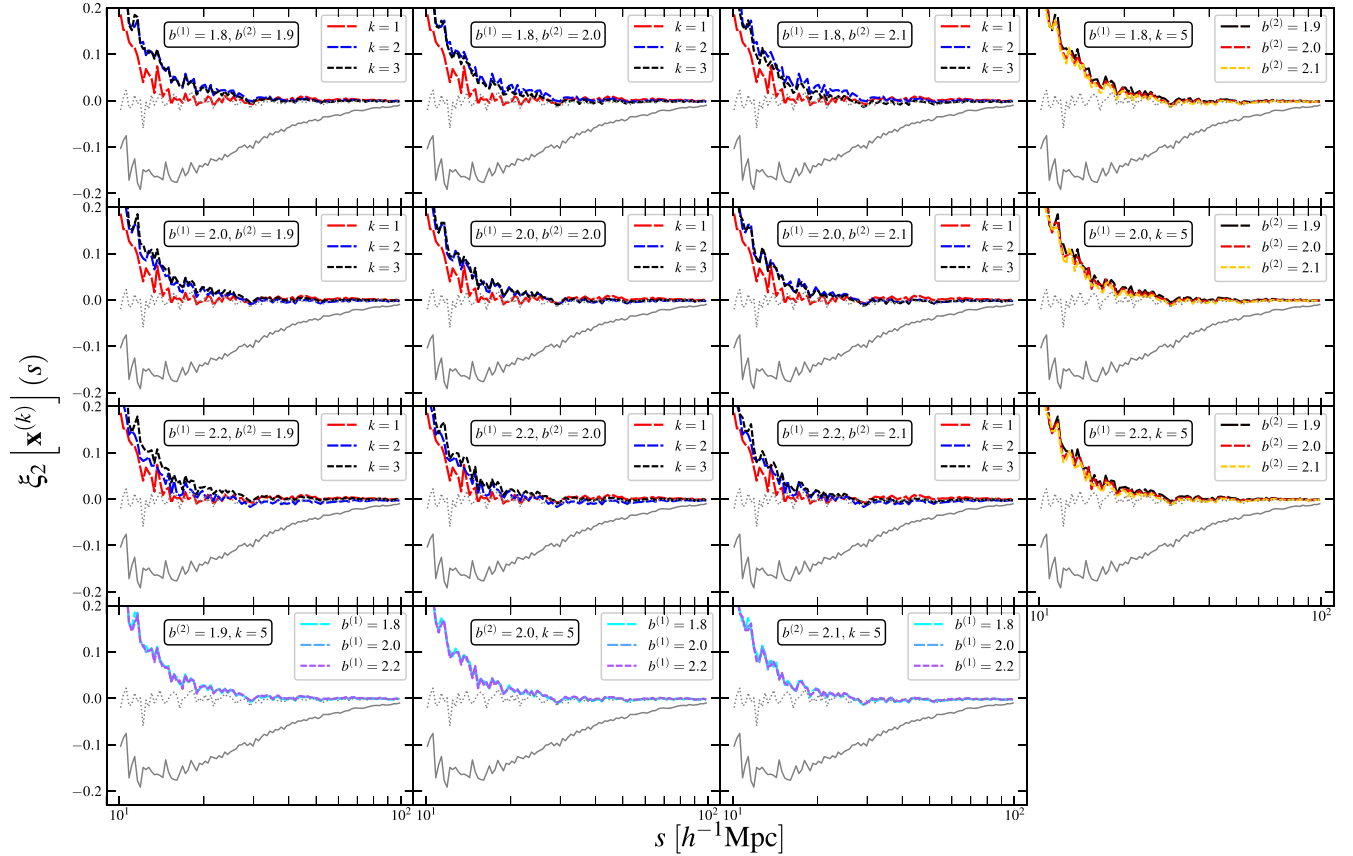
are between the different choices of  $b^{(k)}$ . For all curves in this figure we have fixed  $b^{(0)} = 2.3$  because we have checked that the results for  $b^{(0)} = 2.0$  are almost identical. In the block of  $3 \times 3$  panels at the top left corner, each row has a fixed  $b^{(1)}$  and each column has a fixed  $b^{(2)}$ ; the legend for each curve not only shows the corresponding values of  $b^{(k)}$  but also indicates the current iteration number  $k$ : for example, ‘ $b^{(2)} = 1.9$ ’ in the top left-hand panel means that this is the result after iteration  $k = 3$ , with  $b^{(0)} = 2.3$ ,  $b^{(1)} = 1.8$  and  $b^{(2)} = 1.9$ ,<sup>6</sup> and so on. In all cases, we find that without iterations ( $k = 0$ ) the estimator E3 of the unreconstructed galaxy catalogue is visibly non-zero at  $s \lesssim 60 h^{-1} \text{Mpc}$ , whilst it rapidly converged to 0 at  $s \gtrsim 15 h^{-1} \text{Mpc}$  after one or two iterations. The precise values of  $b^{(1)}$ , in the range of  $[1.8, 2.2]$ , and  $b^{(2)}$ , in the range of  $[1.9, 2.1]$ , have little impact on the final converged results.

Fig. 7 has the same layout as Fig. 6, but shows the results for estimator E4, or  $\xi_0[\mathbf{x}^{(k)}(s)]/\xi_{\text{gg}}(r)$ . This plot again indicates that the exact choices of  $b^{(k)}$  have a relatively small effect, with a larger  $b^{(1)}$  tending to ‘undo’ the improvement by iterative reconstruction (cf. blue curves in the second and third rows), whilst further iterations tending to restore that improvement (black curves in the same panels). In all cases, having  $b^{(k)} = 2.0$  for  $k > 0$  makes E4 close to 1.0, which is

<sup>5</sup>This additional bias is also one of the reasons why the tests do not use the galaxy bias value directly measured from simulations,  $b_{\text{sim}}$ . However, just for completeness, we report the simulation result here –  $b_{\text{sim}} = 1.956$ , which is obtained as the ratio of the galaxy autocorrelation function  $\xi_{\text{gg}}(r)$  and galaxy-matter cross-correlation function  $\xi_{\text{gm}}(r)$  at  $r \gtrsim 5 h^{-1} \text{Mpc}$ , both measured from the simulation using the CUTE code (Alonso 2012).

<sup>6</sup>Note that  $b^{(k)}$  is applied to the galaxy density field  $\delta_{g,s}^{(k)}$  in the  $(k+1)$ th iteration, whilst its effect can be seen in  $\xi_0[\mathbf{x}^{(k+1)}]$ , i.e. after the  $(k+1)$ th iteration, only. Therefore, the curve labelled as ‘ $b^{(2)}$ ’ is actually the result after three iteration. The convention of using  $k$  is indicated in Fig. 1.





**Figure 6.** Estimator E2,  $\xi_2[\mathbf{x}^{(k)}](s)$  for the set of updated galaxy coordinates  $\mathbf{x}^k$  after  $k$ th iterations for different combinations of galaxy bias parameter  $b^{(k)}$ . The bias,  $b^{(k)}$ , is the one used in the  $k+1$ th iteration of the reconstruction method. The values of the  $b^{(1)}$  and  $b^{(2)}$  bias parameters are shown in the label of each plot, whilst  $b^{(0)} = 2.2$  is the same for all panels and thus is not shown. The first three rows contains the tests that use a fixed  $b^{(1)} = 1.8, 2.0$ , and  $2.2$  whilst each column gives the results for a fixed  $b^{(2)} = 1.9, 2.0$ , and  $2.1$ , respectively. So the upper left corner contains  $3 \times 3$  subplots and each of the subplot represents a unique combination of  $b^{(1)}$  and  $b^{(2)}$ , showing the variation of E2 as the iteration number increases. The sparseness of the dashed lines increases with  $k$ ,  $b^{(1)}$  or  $b^{(2)}$ , respectively, in the three different regions (see the legends). The grey dotted and solid lines have the same meaning as in Fig. 4. Each of the subpanels on the rightmost side show how varying  $b^{(2)}$  for a fixed  $b^{(1)}$  affects the reconstruction outcome, whilst each of the subpanels at the bottom illustrate the effect of varying  $b^{(1)}$  for fixed  $b^{(2)}$  values.

not surprising, given that 2.0 is close to the linear bias value measured from the simulations 1.956.

In the left-hand panel of Fig. 8, we present the  $b^{(k)}$  test results for estimator E2. The grey dashed curve at the very bottom is  $r[\delta_g^{(0)}, \delta_g^r]$  or equivalently  $r[\delta_g^{(0)}, \delta_g^r]$ , and the black solid curve immediately above that is  $r[\delta_g^{(1)}, \delta_g^r]$  with  $b^{(0)} = 2.2$ , which indicates that the first iteration substantially improves the correlation between the reconstructed and the real-space galaxy fields. Finally, the top blue curve actually represents a bunch of 12 lines that are so close to each other that they are indistinguishable by eye: these include three lines after the second iteration with  $b^{(1)} = 1.8, 2.0, 2.2$ , and 9 lines after one further iteration, with  $b^{(2)} = 1.9, 2.0, 2.1$ . Together these show that after two iterations the results have converged well. Again, for this estimator we find a weak dependence of the converged result on the values of  $b^{(k)}$ .

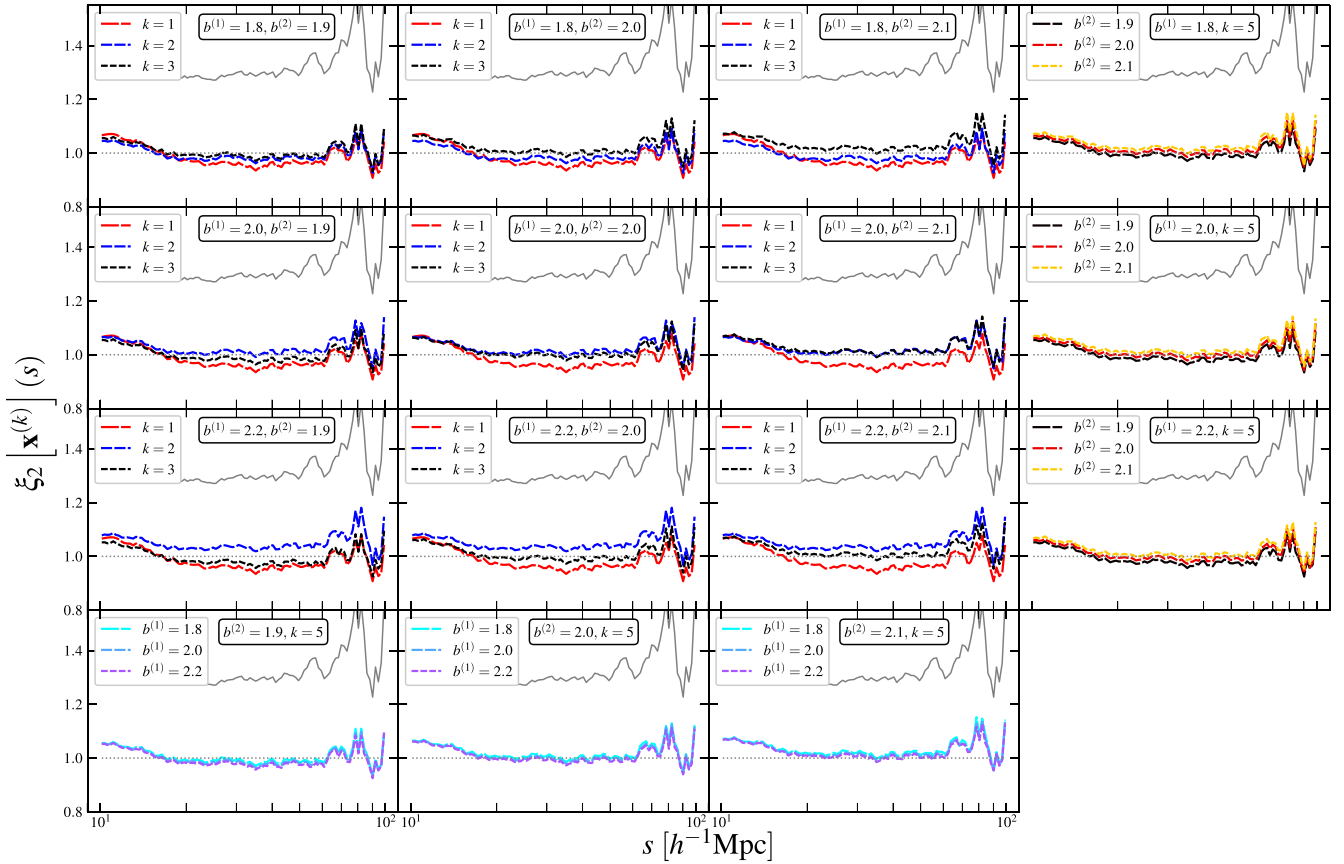
The right-hand panel of Fig. 8 presents the results for E1, in which the black and the grey solid curves are, respectively,  $r[\delta_g^r, \delta_{\text{ini}}]$  and  $r[\delta_g^s, \delta_{\text{ini}}]$  – the cross-correlations between the initial matter density field  $\delta_{\text{ini}}$  and the (non-linear) galaxy density fields from the real- and redshift-space galaxy catalogues (with no iterations in both cases). The purple and grey dashed lines denote, respectively,  $r[\delta_{\text{rec}}, \delta_{\text{ini}}]$  and  $r[\delta_{\text{rec}}^{(0)}, \delta_{\text{ini}}]$  – the cross-correlations between the initial matter density

field and the reconstructed matter density fields from the real- and redshift-space galaxy catalogues (again with no iteration in the latter case). In between the two dashed lines are a bunch of 12 green solid lines – indistinguishable by eye – which show the reconstruction results after three iterations for different combinations of  $b^{(0,1,2)}$ . The iterative RSD reconstruction improves the reconstruction of the initial density field on all scales, whilst there is still some residual RSD effect that it fails to remove.

### 3.3 Linear growth rate $f$

Finally, we have tested the effect of the linear growth rate,  $f$ , in the reconstruction result, using a range of values between 0.5 and 0.9. In our reconstruction algorithm, the size of  $f$  determines how much correction is applied to the redshift-space coordinates of galaxies – a  $f$  value that is too large will make the coordinates overcorrected and vice versa. Therefore, we expect that there is a limited range of  $f$  which would lead to sensible reconstruction result.

We have adopted the following values of the other parameters –  $S = 9 h^{-1} \text{ Mpc}$ ,  $C_{\text{ani}} = 1$ ,  $b^{(0)} = 2.3$ , and  $b^{(k > 0)} = 1.9$  – in all the tests mentioned in this subsection. The left-hand panels of Fig. 9 show the estimator E3,  $\xi_2[\mathbf{x}^{(k)}](s)$ , respectively, for  $f$  from 0.5 to 0.9 (the first



**Figure 7.** The same as Fig. 6, but for estimator E4,  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{gg}(r)$ . The straight dotted and the wiggly solid grey lines are the same in all subpanels: the former is the constant 1.0 to guide the eyes, whilst the latter is the ratio between the multipole moments measured from the redshift- and real-space galaxy catalogues at  $z = 0.5$ .

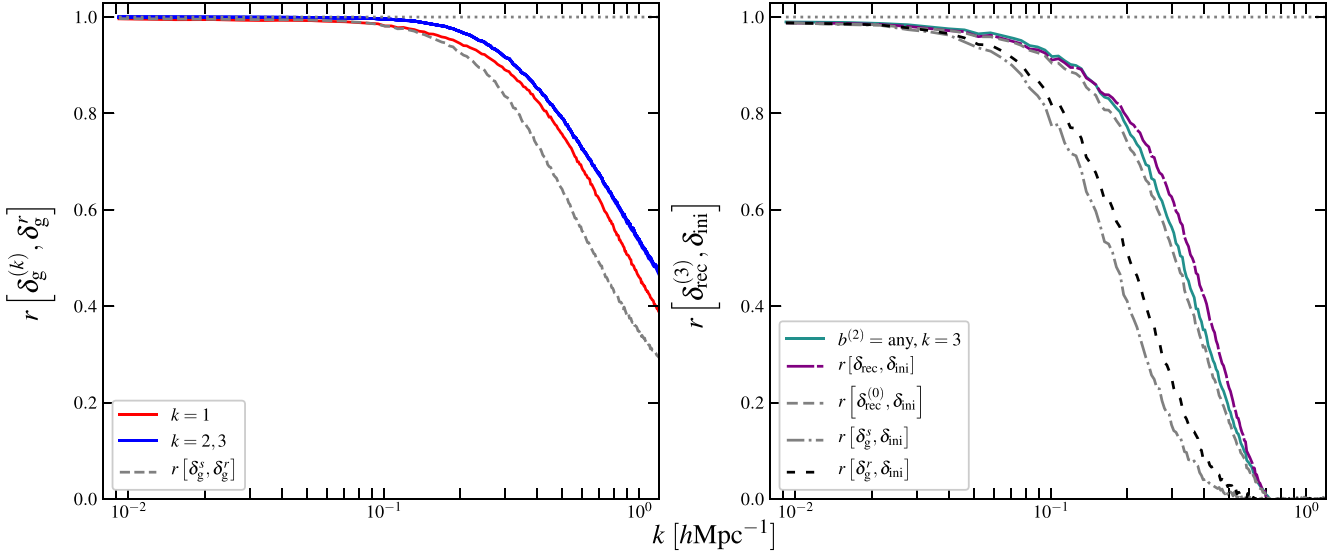
five rows); the last row compares the results from using the different  $f$  values after the fifth iteration. As anticipated above, we confirm that using  $f$  values that are too small ( $f = 0.5, 0.6$ ) leads to incomplete elimination of the quadrupole at  $s \gtrsim 30 h^{-1} \text{ Mpc}$ . Likewise, when the adopted value of  $f$  (e.g.  $f = 0.8, 0.9$ ) is larger than the correct one,  $f = 0.735$ , the quadrupole is overcorrected and becomes slightly positive between 30 and  $40 h^{-1} \text{ Mpc}$ , though both effects are weak.

The right-hand panels of Fig. 9 are the same as the left-hand panels, but for the estimator E4,  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{gg}(r)$ . The behaviour is broadly consistent with what we have found for E3: when  $f$  is too small, the reconstruction, even after convergence, is unable to completely remove the RSD effect and bring E4 to unity, whilst using a value of  $f$  that is too large overcorrects the monopole by making it smaller than the real-space galaxy correlation function.

Their findings seem to suggest that we can use the reconstruction algorithm to place a constraint on  $f$ . However, recall that in linear RSD studies there is a degeneracy between  $f$  and the linear galaxy bias  $b$ . Schematically, the velocity divergence  $\theta$  and linear matter perturbation  $\delta$  are related by  $\theta = aHf\delta$ , whereas the galaxy density field  $\delta_g = b\delta$ , so that  $\theta \propto (f/b)\delta_g \equiv \beta\delta_g$ . Although  $f$  and  $b$  enter the reconstruction pipeline at different places, this degeneracy will persist in the following way: the galaxy density contrast  $\delta_g$  is first divided by  $b$  to get the matter density field; the latter is used to find the displacement field  $\Psi$ ; whilst the reconstruction solves a non-linear equation to calculate  $\Psi$  from  $\delta$ , for large scales the two

quantities satisfy a linear relation to a good approximation, so that  $\Psi$  is ‘proportional’ to  $1/b$ ; then as  $\Psi$  is multiplied by  $f$  according to equation (7) we get the  $f/b$  dependence. As in the tests of this section we have fixed  $b$ , we expect the situation will get more complicated when we allow both  $b$  and  $f$  to be chosen without prior knowledge. Indeed, as we will see in the next section, using estimator E3 alone only gives a constraint on  $\beta$ .

The results for estimators E2 (correlation between the reconstructed and real galaxy density fields) and E1 (correlation between the reconstructed and initial matter density fields) are presented, respectively, in the left-hand and right-hand panels of Fig. 10. Only results after five iterations are shown. The impact of using different values of  $f$  on these estimators is again mild. From the left-hand panel, we can see that for all  $f$  values the correlation is substantially improved compared with the case of no reconstruction (dashed line), and whereas using small values of  $f$  can slightly degrade the performance, using  $f = 0.9$  gives practically indistinguishable result from the case of  $f = 0.735$ . The same happens to estimator E1 which is shown in the right-hand panel, for which the difference between different  $f$  values is even smaller. Overall, we find that, for our chosen galaxy number density,  $n_g = 3.2 \times 10^{-4} [h^{-1} \text{ Mpc}]^{-3}$ , the RSD effect on reconstruction is small, e.g. by comparing  $r[\delta_{\text{rec}}^{(0)}, \delta_{\text{ini}}]$  with  $r[\delta_{\text{rec}}^{(0)}, \delta_{\text{ini}}]$ , and the improvement of the iterations is even smaller, e.g. by comparing the black dotted line with  $r[\delta_{\text{rec}}^{(0)}, \delta_{\text{ini}}]$  (the grey dashed line).



**Figure 8.** *Left-hand panel:* The estimator E2,  $r[\delta_g^{(k)}, \delta_g^r]$  after one (red solid line in the middle), two and three (twelve blue solid lines, which are indistinguishable and appear as the single thick solid line on the top) iterations where the value of  $b^{(k)}$  is allowed to vary across the different iterations; in all cases we have used  $b^{(0)} = 2.2$ ,  $b^{(1)} \in \{1.8, 2.0, 2.2\}$ , and  $b^{(2)} \in \{1.9, 2.0, 2.1\}$ . The blue line is actually 12 overlapping curves, of which 3 have  $k = 2$  and  $b^{(1)} = 1.8, 2.0$  and  $2.2$ , and a further 9 lines which have  $k = 3$  and correspond to all possible combinations of  $b^{(1)}$  and  $b^{(2)}$ . *Right-hand panel:* The same as the left-hand panel, but for estimator E1,  $r[\delta_{\text{rec}}^{(3)}, \delta_{\text{ini}}]$ . Only the results after three iterations are shown (nine green solid lines, which are indistinguishable and appear as the single solid line on the right). The black sparsely dashed (second from the left) and grey dot-dashed (leftmost) lines are the correlation coefficients of the initial matter density field  $\delta_{\text{ini}}$  with the real- and redshift-space galaxy density fields, respectively; the purple long dashed (rightmost) and grey short dashed (left to the solid line) lines are the correlation coefficients of  $\delta_{\text{ini}}$  with the reconstructed matter density field from the real- and redshift-space galaxy density field (no iteration in the latter case), respectively.

#### 4 AN APPLICATION OF THE METHOD

Having now acquired a more complete picture about the behaviour of the reconstruction algorithm and its dependence on the physical and technical parameters, we naturally would like to understand the potential of applying it to real galaxy data to extract cosmological information. Due to the limited scope of this paper, in this section we only attempt to explore this issue for an idealized set-up – reconstruction from mock galaxy catalogues in cubic simulation boxes – and we shall comment on some of the complications when working with real data and leave detailed studies to future works.

As mentioned in the introduction, we wish the reconstruction algorithm to work in a ‘self-calibration’ sense, such that the physical parameters, such as  $f$  and  $b$ , can be determined during the reconstruction process itself. In the previous section, we have seen that it is possible to ‘optimize’ the method using estimator E3,  $\xi_2[\mathbf{x}^{(K)}](s)$ , and also that there is a degeneracy between  $f$  and  $b$  so that it is their combination  $\beta \equiv fb$  that matters. Here, we investigate whether the value  $\beta$  can be accurately fixed by demanding that  $\xi_2[\mathbf{x}^{(K)}](s)$  vanishes for the range of scales in which the iterative reconstruction method works.

In linear perturbation theory, the integration of Kaiser formula (Kaiser 1987) to relate the real- and redshift-space galaxy density contrasts,

$$\delta_g^s(s, \mu) = (1 + \beta \mu^2) \delta_g^r(r), \quad (18)$$

where  $\beta$  is the distortion parameter introduced above, and  $\mu$  is the cosine of the angle between the galaxy pair separation vector  $\mathbf{s}$  and the LOS direction, gives the monopole and quadrupole moments of the galaxy correlation function in redshift space (Hamilton 1993)

$$\xi_0(s) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) \xi_{\text{gg}}(r), \quad (19)$$

$$\xi_2(s) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) [\xi_{\text{gg}}(r) - \bar{\xi}_{\text{gg}}(r)], \quad (20)$$

where  $\bar{\xi}_{\text{gg}}(r)$  is the average correlation function with  $r$ :

$$\bar{\xi}_{\text{gg}}(r) \equiv 3r^{-3} \int_0^r \xi_{\text{gg}}(r') r'^2 dr'. \quad (21)$$

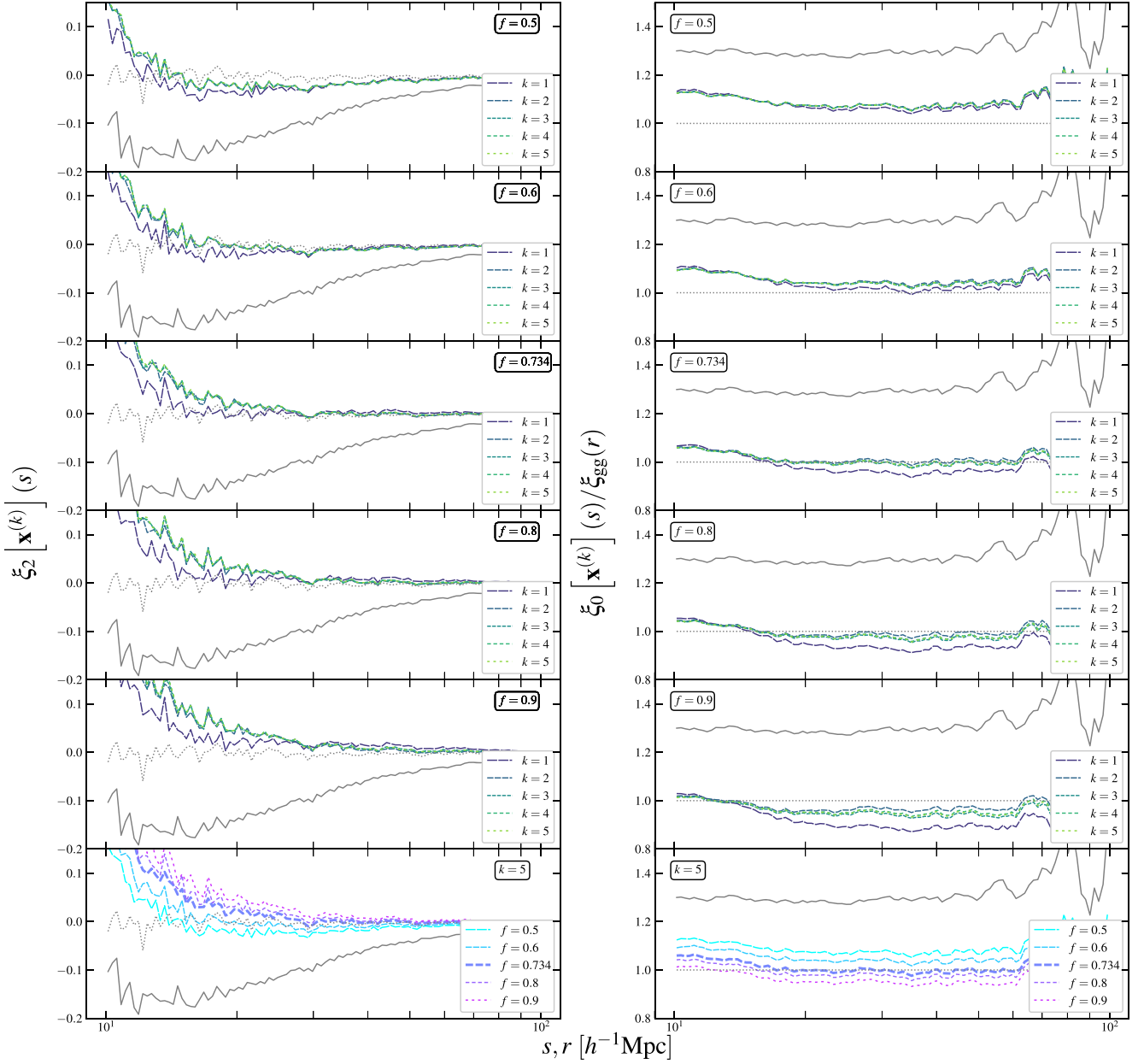
In practice, whilst  $\xi_{0,2}$  can be directly measured from observations, these equations cannot be used to infer  $\beta$  because the real-space correlation function  $\xi_{\text{gg}}(r)$  is not observable. However, it is possible to combine these equations to estimate  $\beta$  from the ratio of redshift-space quadrupole and monopole moments:

$$\frac{\xi_2(s)}{\xi_0(s) - \bar{\xi}_0(s)} = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2}, \quad (22)$$

with

$$\bar{\xi}_0(s) \equiv 3s^{-3} \int_0^s \xi_0(s') s'^2 ds'. \quad (23)$$

The method has been applied to galaxy redshift surveys (Peacock et al. 2001; Hawkins et al. 2003), but studies using simulation data show that it is difficult to accurately recover the actual value of  $\beta$  (e.g. Cole et al. 1994; Hernández-Aguayo et al. 2019), whilst the estimator  $\xi_0(s)/\xi_{\text{gg}}(r)$  gives better measurement of  $\beta$ . One reason for this is that the estimator given in equation (22) suffers from the often noisier measurement of  $\xi_2(s)$ . Furthermore, whilst Kaiser effect only takes account for the coherent linear motion of galaxies, the non-linearity induced by FoG effect could have a non-negligible impact on the quadrupole down to  $s \sim 50 h^{-1} \text{ Mpc}$  (Cole et al. 1994), which



**Figure 9.** *Left-hand panels:* The estimator E3,  $\xi_2[\mathbf{x}^{(k)}](s)$  for  $f = 0.5, 0.6, 0.735$  (theoretical value),  $0.8$  and  $0.9$  in the first five rows after  $k$ th iteration, respectively, plotted in dashed lines with their sparsenesses increasing with  $k$  (see the legends). The grey dotted curve (close to constant zero) is the quadrupole measured from the real-space galaxy catalogue, whilst the grey solid curve is that measured from the redshift-space catalogue before reconstruction. The last row compares the results from using the different  $f$  values after the fifth iteration, with the sparsenesses of the curves increasing with  $f$  (see the legends). *Right-hand panels:* The same as the left, but for estimator E4,  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{\text{gg}}(r)$ . The grey solid curve is  $\xi_0(s)/\xi_{\text{gg}}(r)$ , i.e. result before reconstruction, and the horizontal grey dotted line is 1.

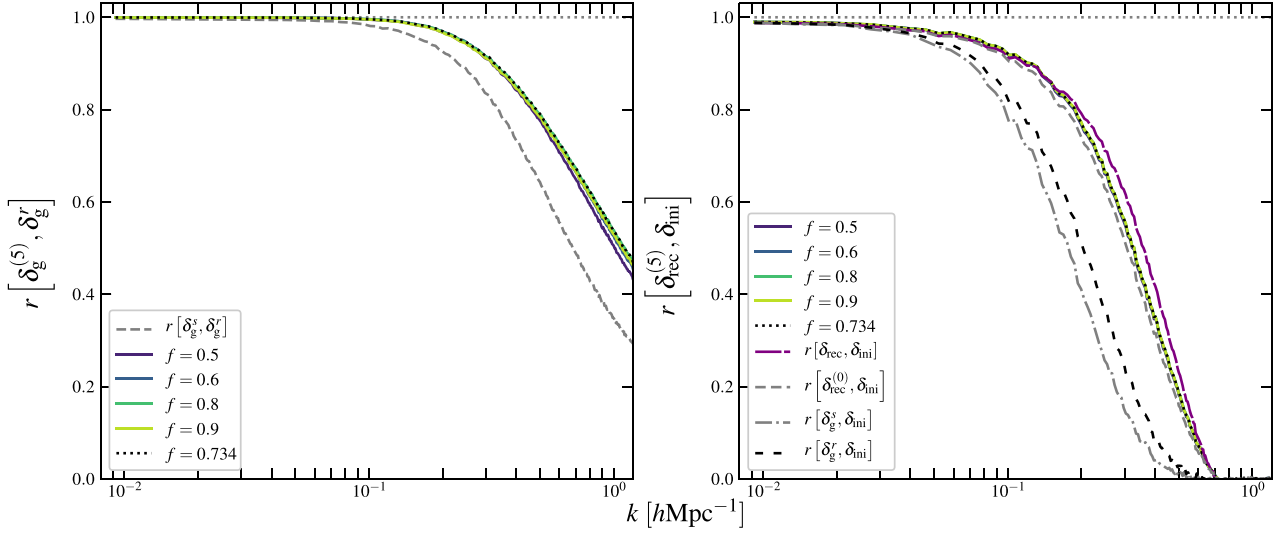
invalidates the linear assumption in equation (22). Thus, using only the monopole moment in extracting information according to linear theory could be desirable over the traditional method equation (22): even though this requires knowledge of  $\xi_{\text{gg}}(r)$  which is unobservable, we have seen that reconstruction gives  $\xi_0[\mathbf{x}^{(k)}](s)$  that can be used as an approximation to  $\xi_{\text{gg}}(r)$ .

In order to more quantitatively assess the accuracy of this approximation, we have measured  $\xi_{2,0}[\mathbf{x}^{(k)}](s)$  from the 15 realizations of mock galaxy catalogues, and the results are shown, respectively, in the upper left and upper middle panels of Fig. 11. There we have adopted the theoretical values of  $f = 0.735$  and  $b = 1.95$ , so that

$\beta = f/b = 0.376$ , for reconstructions in all 15 realizations; the thin lines are the results from the individual realizations whilst the thick lines are their means. We can see that using the ‘correct’ values of  $f$ ,  $b$  gives  $\xi_2[\mathbf{x}^{(k)}](s) = 0$  and  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{\text{gg}}(r) = 1$  at  $s \gtrsim 20$ , with an accuracy of  $\sim 0.01$  and  $\sim 1$  per cent, respectively. This confirms that we can actually replace  $\xi_{\text{gg}}$  in equation (19) with  $\xi_0[\mathbf{x}^{(k)}]$  as a way to estimate  $\beta$ :

$$R(s) \equiv \frac{\xi_0(s)}{\xi_0[\mathbf{x}^{(k)}](s)} = 1 + \frac{2}{3}\beta(s) + \frac{1}{5}\beta^2(s). \quad (24)$$





**Figure 10.** *Left-hand panel:* Estimator E2,  $r[\delta_g^{(k)}, \delta_g^r]$  (the correlation between the reconstructed and real-space galaxy density fields), for different values of  $f$  after five iterations of the reconstruction procedure. The black dotted line represents  $f = 0.734$  and other  $f$  values are shown by coloured solid lines as indicated in the legend (note that all the coloured solid lines but the case for  $f = 0.5$  – which is the lowest of them – are indistinguishable). The grey dashed curve is the correlation between the redshift-space galaxy density field before reconstruction,  $\delta_g^{(0)}$ , and the real-space galaxy density field  $\delta_g^r$ . *Right-hand panel:* The same as the left, but for estimator E1,  $r[\delta_{\text{rec}}^{(k)}, \delta_{\text{ini}}]$  (the correlation between the reconstructed and initial matter density fields). The black sparsely dashed (second from the left) and grey dot-dashed (leftmost) lines are the correlation coefficients of the initial matter density field  $\delta_{\text{ini}}$  with the real- and redshift-space galaxy density fields, respectively; the purple long dashed (rightmost) and grey dashed (third from the left) lines are the correlation coefficients of  $\delta_{\text{ini}}$  with the reconstructed matter density field from the real- and redshift-space galaxy density field (no iteration in the latter case), respectively. Note that all the coloured solid lines (for different  $f$  values) are indistinguishable from each other, and the black dotted line, which is for  $f = 0.734$ , is on top of them [they appear as a single line (second from the right)].

Note that in linear theory  $\beta$  is a constant, but here we have kept the  $s$ -dependence because the estimator itself can fluctuate around the constant value from one  $s$  bin to another.

To use the estimator  $R(s)$ , we must first specify the range of  $s$ ,  $[s_{\min}, s_{\max}]$ , within which  $R(s)$  is evaluated. We choose  $s_{\min}$ ,  $s_{\max}$  to ensure that  $R(s)$  is approximately constant in that range, which gives  $s_{\min} \simeq 22 h^{-1} \text{ Mpc}$ ; for  $s_{\max}$ , we have checked three different values,  $71.6 h^{-1} \text{ Mpc}$ ,  $90.2 h^{-1} \text{ Mpc}$ , and  $98.9 h^{-1} \text{ Mpc}$ : in principle, we expect that  $R(s)$  is closer to a constant at larger  $s$ , but in practice the estimator becomes quite noisy and the uncertainty becomes large in that regime due to cosmic variance, so that we expect including larger  $s$  bins should have a relative small impact on the overall fitting result. We divide  $[s_{\min}, s_{\max}]$  into  $N = 10$  bins equally spaced in logarithmic scale and use a least chi-square method to find the best-fitting by minimizing  $\chi^2$

$$\chi^2 = \frac{1}{N-1} \sum_{i=1}^N \frac{(R_i - R)^2}{\sigma_i^2}, \quad (25)$$

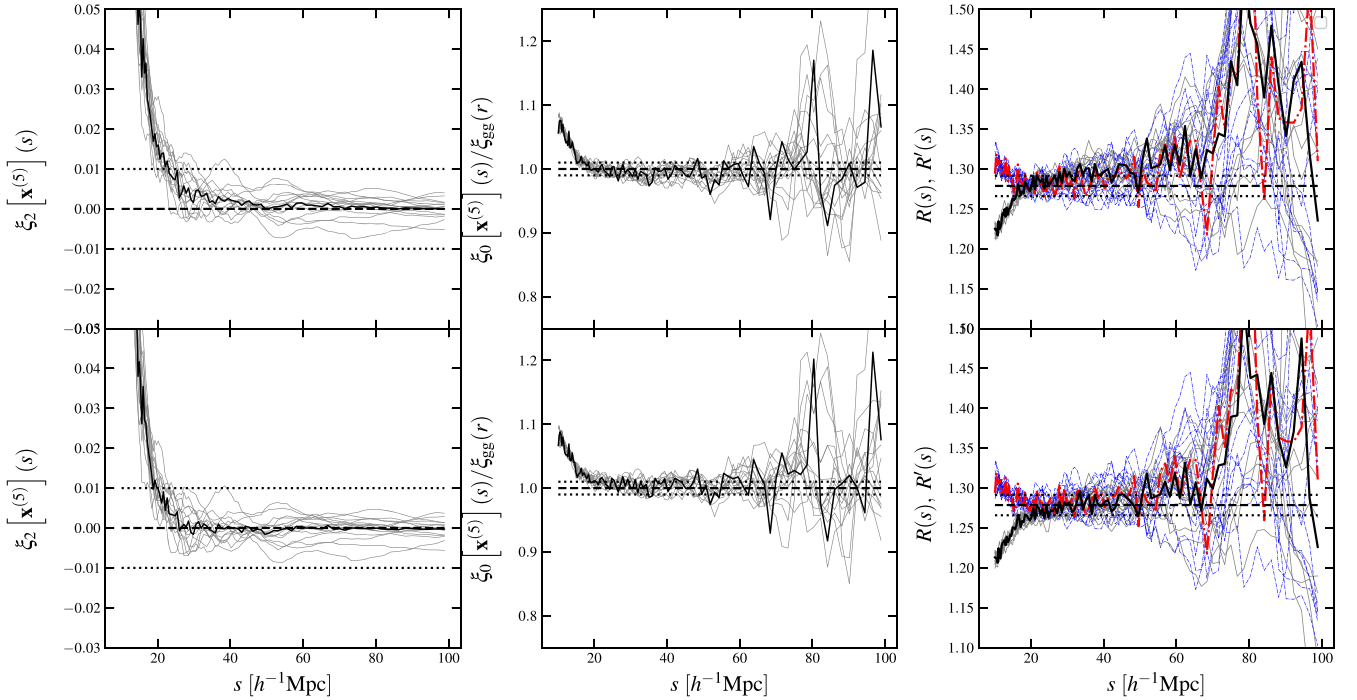
where  $R \equiv 1 + \frac{2}{3}\beta_{\text{BF}} + \frac{1}{5}\beta_{\text{BF}}^2$  with  $\beta_{\text{BF}}$  the best-fitting value of  $\beta$ ,  $R_i$  is the average value of  $R(s)$  in the  $i$ th bin, and  $\sigma_i$  is the corresponding standard deviation of the 15 realizations.

In the upper right-hand panel of Fig. 11, we have compared the estimator in equation (24) with the estimator  $R'(s) \equiv \xi_0(s)/\xi_{\text{gg}}(r)$  from equation (19), in which the thin blue (grey) lines denote the latter (former) for the 15 individual realizations, whilst the thick red (black) line denotes the mean of them. This plot shows that at  $s \gtrsim 20 h^{-1} \text{ Mpc}$   $R(s)$  and  $R'(s)$  agree with each other at percent level. We have also checked the inferred  $\beta_{\text{BF}}$  values, following the method described in the previous paragraph, from these two estimators respectively, and found a good agreement too. Note that a

small error in the estimator  $R(s)$  or  $R'(s)$  can be translated to a larger error in  $\beta$  because of the  $1$  in  $1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2$ .

It is also interesting to note in the upper right-hand panel of Fig. 11 that the reconstruction result  $R(s)$  (thick black line) is slightly less noisier than the theoretical result  $R'(s)$  (thick red line). The noise in  $R'(s)$  is generated because the random phases of the galaxy field are changed by the RSD mapping from real space ( $\delta_g^r(\mathbf{r})$ ) to redshift space ( $\delta_g^s(\mathbf{s})$ ) – i.e.  $\delta_g^s(\mathbf{s})$  is not simply a constant amplitude enhancement of  $\delta_g^r(\mathbf{r})$  – so that when taking the ratio  $\xi_0(s)/\xi_{\text{gg}}(r)$  there is no perfect cancellation. The reconstruction works to revert this change of phases, but in practice this cannot be done perfectly, which means that, compared with the real-space galaxy field, the phases of the reconstructed galaxy field are ‘more similar’ to those of the redshift-space galaxy field. Thus, by dividing the two function as in equation (24), there is a better cancellation of the phase effect.

In practice, we do not know the theoretical values of  $f$  and  $b$  a priori, and it is interesting to know what happens to the inferred  $\beta$  value if incorrect values of  $f$ ,  $b$  are used in the reconstruction. In Birkin et al. (2019), for example, it was found that the reconstruction result of the initial condition phases is not particularly sensitive to the value of the linear bias  $b$  used. Here, we have checked this by running the reconstruction pipeline on the 15 realizations of mock galaxy catalogues, using 7 different  $(f, b)$  combinations, including two which give very similar values of  $\beta = fb$ . More explicitly, we have adopted a  $3 \times 2$  grid with  $f \in \{0.694, 0.712, 0.734\}$ ,  $b \in \{1.96, 2.0\}$ , and two points with  $(f, b) = (0.75, 1.92)$  and  $(0.734, 1.956)$  to enrich the result. The corresponding  $\beta$  ranges from 0.347 to 0.391 passing through 0.354, 0.356, 0.363, 0.367, 0.374 (theoretical value), and 0.375. In these tests, the other reconstruction parameters are fixed to  $S = 9 h^{-1} \text{ Mpc}$  and  $C_{\text{ani}} = 1.0$ .



**Figure 11.** The quadrupole ( $\xi_2[\mathbf{x}^{(s)}](s)$ ; left column), monopole ( $\xi_0[\mathbf{x}^{(s)}](s)/\xi_{\text{gg}}(r)$ ; central column), and  $R(s) \equiv \xi_0(s)/\xi_0[\mathbf{x}^{(s)}](s)$  (right column) of the reconstructed galaxy catalogues from the 15 realizations of mock redshift-space galaxy catalogues. The results for individual realizations are plotted as thin grey solid curves with their mean as thick black solid curves. In the right column, we also show the results of  $R'(s) \equiv \xi_0(s)/\xi_{\text{gg}}(r)$ , the ratio between the redshift-space monopole and the real-space galaxy correlation function for comparison, with the thin blue dash-dotted lines showing the individual realizations and the thick red dash-dotted lines their mean. The top row shows the reconstruction result using theoretical values  $f = 0.734$  and  $b = 1.95$ , whilst the bottom row shows the result using the best-fitting value of  $\beta = f/b$  obtained by minimizing the derivation of  $\xi_2[\mathbf{x}^{(s)}](s)$  from 0 (see Section 4 for details), which corresponds to  $\beta = 0.356$ . The dotted horizontal lines in the left and central columns mark the  $\pm 0.01$  deviation from 0 and 1, and in the right column they mark the  $\pm 1$  per cent deviation from  $R(s)$  calculated according to equation (24) with the theoretical value of  $\beta$ .

From the reconstructed catalogues we then measure  $R(s)$  and use them to find the best-fitting  $\beta_{\text{BF}}$ . In the upper left-hand panel of Fig. 12, we plot  $\beta_{\text{BF}}$  against  $f/b$  for these points as crosses; for each  $f/b$  value, there are three  $\beta_{\text{BF}}$  values with different colours, respectively, obtained by using  $s_{\text{max}} = 98.9 h^{-1} \text{Mpc}$  (green),  $90.2 h^{-1} \text{Mpc}$  (red), and  $71.6 h^{-1} \text{Mpc}$  (blue). We make the following observations from this plot:

(i)  $\beta_{\text{BF}}$  versus  $f/b$  falls nicely on a straight line with slope 1 (the solid line). However, note that the relation  $\beta_{\text{BF}} = f/b$  (the dashed line) does not exactly hold: we have checked that the actual value of  $\beta_{\text{BF}}$  depends on the  $s_{\text{min}}$  used, decreasing as smaller  $s_{\text{min}}$  is used; the results here are for  $s_{\text{min}} = 22.1 h^{-1} \text{Mpc}$ . The trend suggests that, roughly speaking, the reconstruction simply spits out the ‘input’  $\beta$  value, and therefore has no predictive power regarding  $\beta$ , unless we look beyond  $R(s)$ .

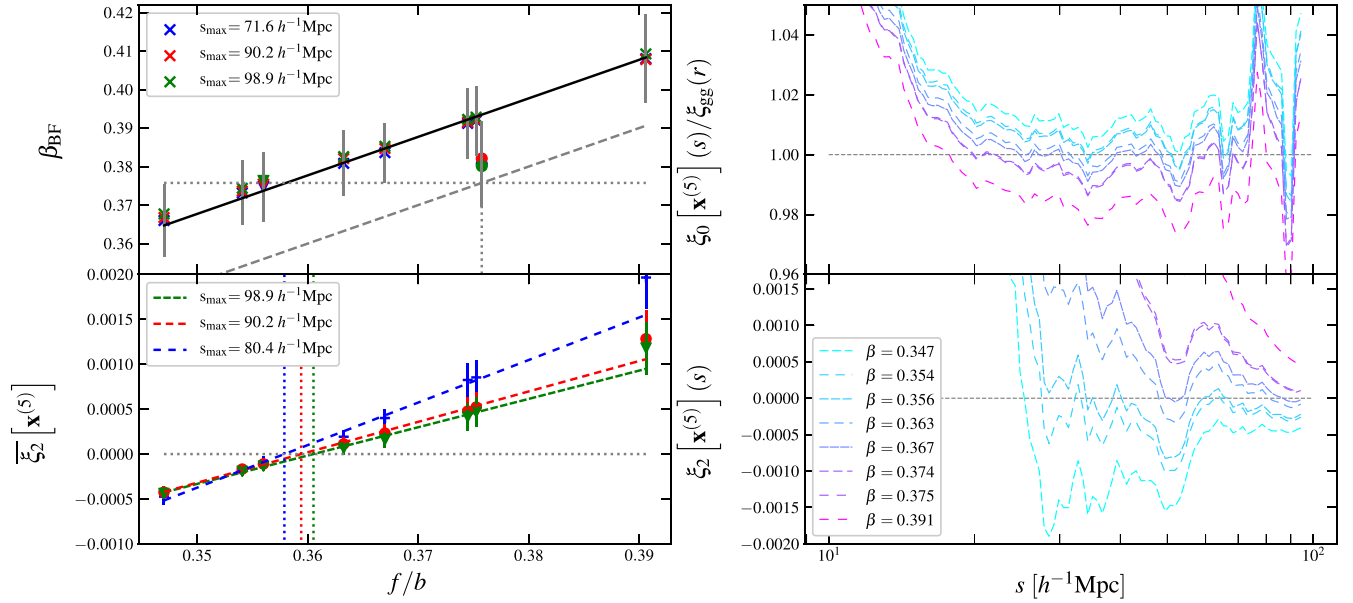
(ii) The two  $(f, b)$  pairs that give similar  $\beta = f/b \approx 0.375$  produce very similar  $\beta_{\text{BF}}$ : this is a consequence of the  $f$ - $b$  degeneracy mentioned above.

(iii) For comparison, we include the best-fitting  $\beta_{\text{BF}}$  values obtained using  $R'(s)$  as filled circles, using the same colour scheme. Because there is no reconstruction and therefore no input  $(f, b)$ , we simply plot the points at a fixed horizontal coordinate  $f/b = 0.376$  (the theoretical value). We note that  $\beta_{\text{BF}}$  obtained in this way is indeed  $\sim 1$  per cent larger than the theoretical value, but this is well within the  $1\sigma$  uncertainty indicated by the grey error bar; at this stage, we cannot accurately assess the impact on  $\beta_{\text{BF}}$  due to the small sample size or the simulation resolution, which is beyond the scope of this

work. As mentioned above, a small (per cent-level) error in  $R'(s)$  leads to a larger error in  $\beta_{\text{BF}}$ .

This has motivated us to consider an alternative approach to find the best-fitting value of  $\beta$  from the reconstruction process, by demanding that  $\xi_2[\mathbf{x}^{(K)}](s)$  is closest to zero on large, linear, scales. In practice, we quantify this using the mean value of  $\xi_2[\mathbf{x}^{(K)}](s)$  between  $s'_{\text{min}} \simeq 40 h^{-1} \text{Mpc}$  and  $s_{\text{max}} = 98.9, 90.2, 80.4 h^{-1} \text{Mpc}$ ,  $\bar{\xi}[\mathbf{x}^{(K)}]$ . In the lower left-hand panel of Fig. 12, we present  $\bar{\xi}[\mathbf{x}^{(K)}]$  for the seven  $(f, b)$  pairs as used above; the data points with different colours indicate the three choices of  $s_{\text{max}}$  and the dashed lines are the best-fitting straight lines through the points. We note that as  $f/b$  increases  $\bar{\xi}[\mathbf{x}^{(K)}]$  goes from negative to positive, crossing zero at  $f/b \approx 0.36$ . This is indeed  $\sim 4$  per cent smaller than the theoretical value  $f/b = 0.376$ , but note that the resulting  $\beta_{\text{BF}}$  corresponding to this  $f/b$  value (inferred from the upper left panel of Fig. 12) is actually closer to the best-fitting value  $\beta_{\text{BF}} \approx 0.38$  from  $R'(s)$ . This result is not surprising given that, as we shall see shortly, when using input value  $f/b \simeq 0.36$  the E4 estimator is very close to 1.0 and  $R(s)$  agrees better with  $R'(s)$  at  $s \gtrsim 20 h^{-1} \text{Mpc}$  than the case where  $f/b$  takes the theoretical value 0.376. We suspect that this small discrepancy is again due to the limited sample size or simulation resolution, and a more detailed investigation into this issue will be conducted in a forthcoming work with larger data sets.

The results of  $\xi_2[\mathbf{x}^{(K)}](s)$  and of  $\xi_0[\mathbf{x}^{(K)}](s)/\xi_{\text{gg}}(r)$  for the seven  $(f, b)$  pairs are respectively displayed in the upper right and lower right-hand panels of Fig. 12. The  $(f, b)$  pair whose corresponding  $\beta$  value is closest to 0.38 is  $f/b = 0.356$  (the third point from the left



**Figure 12.** *Upper left:* The best-fitting  $\beta_{\text{BF}}$  values obtained using the estimator  $R(s)$  and the reconstructed galaxy catalogues, for seven choices of input parameters  $f$  and  $b$  supplied to the reconstruction pipeline (crosses).  $R(s)$  is measured as the weighted average between  $s_{\text{min}} = 22.1 h^{-1} \text{ Mpc}$  and  $s_{\text{max}} = 98.9$  (green),  $90.2$  (red), and  $71.6 h^{-1} \text{ Mpc}$  (blue), respectively, and the error bars are the standard deviations obtained from 15 realizations for the case of  $s_{\text{max}} = 71.6 h^{-1} \text{ Mpc}$  as the error bars for remaining  $s_{\text{max}}$  values are similar. The filled circles are the same, but obtained from the estimator  $R'(s)$ : since there is no reconstruction in this case, these data are plotted against the theoretical value  $\beta = f/b = 0.376$ . The dashed line represents  $\beta = f/b$ , whilst the solid line is a line of slope 1.0 to guide the eyes. *Lower left:*  $\xi_2[\mathbf{x}^{(k)}]$ , the constant value fitting  $\xi_2[\mathbf{x}^{(k)}](s)$  between  $s'_{\text{min}} = 39.4 h^{-1} \text{ Mpc}$  and the three  $s_{\text{max}}$  values  $s_{\text{max}} = 98.9$  (green, densely dashed),  $90.2$  (red, dashed), and  $80.4 h^{-1} \text{ Mpc}$  (blue, sparsely dashed), respectively, for the same seven choices of input  $(f, b)$ . The vertical dotted lines mark the best-fitting  $\beta_{\text{BF}}$  at which  $\xi_2[\mathbf{x}^{(k)}]$  is closest to zero, and the error bars are again the standard deviations of the 15 realizations. *Right-hand panels:* the estimator E4 (upper) and E3 (lower) for the seven choices of reconstruction parameters  $(f, b)$ , with their corresponding values of  $\beta$  indicated in the legend and increase with the sparsenesses of the curves [ $\beta$  increases from top (bottom) to the bottom (top) line in the upper (lower) right-hand panels]. Note that according to the lower left-hand panel, the best value of  $\beta$ , amongst the seven choices, is  $\beta = 0.356$ , and results for this choice are shown as triangles in the upper left-hand panel.

in the upper/lower left-hand panels of Fig. 12), and we can see that for this value  $\xi_2[\mathbf{x}^{(k)}](s)$  is actually closest to zero (as expected) and  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{\text{gg}}(r)$  closest to 1.0. This latter observation in particular indicates that by demanding  $\xi[\mathbf{x}^{(k)}]$  to be closest to zero for the best-fitting  $\beta$ , we automatically obtain the  $\xi_0[\mathbf{x}^{(k)}](s)$  that is closest to  $\xi_{\text{gg}}(r)$ .

To compare the performance of this new method to determine  $\beta_{\text{BF}}$ , in the lower panels of Fig. 11 we have shown the same curves as in the upper panels, but this time  $\beta = f/b = 0.356$ . A careful visual inspection shows that the results of  $\xi_2[\mathbf{x}^{(k)}](s)$  (lower left-hand panel) and of  $\xi_0[\mathbf{x}^{(k)}](s)/\xi_{\text{gg}}(r)$  (lower middle panel) are actually closer to 0.0 and 1.0, respectively, than the cases shown in the upper panels of Fig. 11, where the theoretical value  $f/b = 0.376$  is used. In the lower right-hand panel, we again find that  $R(s)$  (black thick line) with  $f/b = 0.356$  is nearly identical to  $R'(s)$  (red thick line, which is the same as in the upper right-hand panel) for  $s \gtrsim 20 h^{-1} \text{ Mpc}$ . This is why in the above we have adopted  $s_{\text{min}} \simeq 22 h^{-1} \text{ Mpc}$  for constraining  $\beta_{\text{BF}}$  using  $R(s)$  and  $R'(s)$ .

As highlighted above, in this section we have proposed to obtain the best-fitting value of  $\beta$  using a ‘self-calibration’ of the observed (or mock) galaxy field, namely tune  $\beta$  to minimize  $\xi_2[\mathbf{x}^{(k)}]$ . From lower left-hand and upper right-hand panels of Fig. 12, we have seen that the quadrupole of the reconstructed galaxy field changes slowly with  $f/b$ . The reconstruction method is a backward modelling approach, where one starts with a late-time observed galaxy field to infer information about the early-time density field, the peculiar velocity field and the cosmological model ( $\beta$ ). This is different from

the standard forward modelling approach where one starts with a given cosmological model, makes prediction and compares it against observations (see e.g. Sanchez et al. 2017). It is possible to adopt a hybrid approach. For example, one can start with a specific model, e.g.  $\Omega_m, \sigma_8$ , and predict the linear growth rate  $f$  and linear matter correlation function at a  $z$ . The latter, together with  $\xi_0[\mathbf{x}^{(k)}](s)$  as an approximation to  $\xi_{\text{gg}}(r)$ , can be used to determine the linear bias  $b$  and break its degeneracy with  $f$ . The former can then be checked against the  $f$  determined as the ratio between the best-fitting  $\beta_{\text{BF}}$  and the  $b$  obtained in this way, and against the velocity power spectrum from the reconstructed galaxy catalogue. A detailed investigation of this, however, is beyond the scope of this work.

## 5 DISCUSSION AND CONCLUSIONS

We have proposed and tested a new iterative scheme to reconstruct, simultaneously, the real-space *galaxy coordinates* at late times and the *initial matter density field*, from a given late-time redshift-space galaxy catalogue. The method builds on the non-linear reconstruction algorithm developed by Shi et al. (2018), Birkin et al. (2019), taking into account the (linear) galaxy bias  $b$  and RSDs. It employs a number of (semi)-free parameters related to the smoothing of the galaxy field,  $b$  and linear growth rate  $f$ . Since this is a new method and code, we have performed various checks not only to assess its reliability in obtaining physical results but also to understand how the reconstruction outcomes react to the specific choices of technical parameters.

The iterative reconstruction consists of continuous updates of the trial real-space galaxy coordinates based on improved knowledge of the large-scale displacement field  $\Psi$ , until the results, as quantified by estimators E1–E4,<sup>7</sup> converge. We have found that for reasonable choices of parameters – mainly the smoothing length  $S$  – convergence can be achieved quickly, often after  $2 \sim 4$  iterations. The final results are fairly insensitive to the galaxy bias parameters after three iterations, but depend more strongly on the smoothing parameters  $S$ ,  $C_{\text{ani}}$  and the linear growth rate  $f$  (particularly when  $f$  is too small). With our optimal choice of  $S$ ,  $C_{\text{ani}}$ ,  $b^{(k)}$ ,  $f$ , we find that the method can accurately eliminate the quadrupole moment of the correlation function of the reconstructed galaxy catalogue and reproduce the monopole at  $s \gtrsim 20 h^{-1}$  Mpc. One thing worthwhile to note here is that the elimination of galaxy correlation quadrupole by tuning  $S$  and  $C_{\text{ani}}$  can be done on real observational data without the use of mock catalogues, which means that the method (with appropriate generalization to include effects such as survey geometry and completeness) can be applied directly to real data. This can be considered as some sort of *self* or *internal* calibration of reconstruction parameters, which may help us to avoid relying too heavily on mock data for guiding the choice of parameters.

We can quantify the resulting improvement of the reconstruction method by measuring  $k_{90}$  and  $k_{50}$  which are, respectively, the values of  $k$  where the correlation coefficient between two distribution drops below 90 per cent and 50 per cent. In Table 2, we summarize the values read from Fig. 8. The result of  $k_{50}$  shows that our iterative RSD removal method increases the  $k$ -range that has a good (i.e. higher than 50 per cent) correlation between the reconstructed and the true real-space galaxy positions by 50 per cent. The improvement is even stronger when estimating the initial conditions, with an 80 per cent increase of  $k_{50}$  in the correlation between the reconstructed matter density and the true linear density field, compared with the case in which no reconstruction has been applied, if the reconstruction is performed on the redshift-space galaxy distribution. This is similar to, albeit slightly lower than, the case when the initial condition is reconstructed from the real-space galaxy distribution; in this latter case  $k_{50}$  increases by a factor of 2. This confirms that redshift-space distortions have a negative impact that cannot be completely undone by this iterative RSD removal method, but the latter leads to a substantial gain.

The galaxy catalogue we used in this work for tests has a relatively low number density,  $n_g \simeq 3.2 \times 10^{-4} [h^{-1} \text{Mpc}]^{-3}$ , as a result of which the RSD effect on the reconstruction of initial matter density field is small (e.g. by comparing the dashed grey lines in the right-hand panel of Fig. 8 or Fig. 10). Nevertheless, correcting for RSD still improves the reconstruction result, by increasing the correlation of the reconstructed density field with the true initial density field of our test simulation. It will be interesting to analyse how the method works for galaxy catalogues with higher number density or with other tracers (such as 21cm intensity maps or quasars), in particular whether different choices of parameters need to be made in those cases; we will leave a more detailed investigation of these issues into a future publication.

The observations that the quadrupole moment can be successfully brought back to 0.0, and the monopole close to the real-space

galaxy correlation function, by the iterative reconstruction method at  $s \gtrsim 20 h^{-1}$  Mpc, raise the interesting question whether the method here has done more than removing linear RSD on these scales. This question would be best answered by quantifying how the constraints on cosmological parameters can be improved by going from linear RSD modelling to analysing summary statistics that are extracted from a reconstructed galaxy catalogue. This is beyond the scope of this work, not least because the current pipeline still needs further extensions to account for various observational systematics (as mentioned below), but also because future work is needed to understand the covariance after reconstruction (which is non-trivial given that cosmological parameters are used in both the reconstruction-based RSD removal itself and the theoretical modelling of said summary statistics). Without such an analysis available, a somehow indirect way to gain some insight into this question is to look at the scale at which non-linearity already needs to be accounted for when theoretically predicting the redshift-space correlation function quadrupole. As an example, Cuesta-Lazaro et al. (2020, fig. 14) compared the quadrupoles measured from  $N$ -body simulations and from a Gaussian streaming model where the real-space correlation function and galaxy pairwise velocity moments were obtained using the convolutional Lagrangian perturbation theory, and found the latter deviates from simulation results already at scales as large as  $40\text{--}50 h^{-1}$  Mpc. We also know that the linear Kaiser model does not match simulation data well at scales  $\lesssim 40\text{--}50 h^{-1}$  Mpc (see e.g. the right-hand panel of fig. 5 in Hernández-Aguayo et al. 2019, which used the same set of GR simulations as in this paper). These imply that the iterative method does more than linear RSD removal – which is not surprising given that, unlike the standard Zel’dovich approximation, here the reconstruction method gives a non-linear displacement field.

The results also imply that it might be possible to use this reconstruction method to infer statistical information about the large-scale peculiar velocity field. The exact details of this information, the ranges of scale and velocity within which it can be reliably extracted, as well as the accuracy of the results, are again beyond the scope of a single paper. Velocity reconstruction has been studied by several groups (e.g. Wang et al. 2012; Yu & Zhu 2019) in a similar context, but with different reconstruction algorithms, including the one used here but with no iterations. It will be of interest to test the iterative reconstruction method presented in this paper for galaxy catalogues with different number densities and at different redshifts in a future work.

We checked the possibility of using this reconstruction method to determine the value of  $\beta$ , and found that the best-fitting  $\beta$  from the estimator  $R(s) = \xi_0(s)/\xi_0[\mathbf{x}^{(k)}](s)$  follows the  $\beta$  corresponding to the input values of  $f$ ,  $b$  in the reconstruction, and therefore  $R(s)$  alone cannot fix  $\beta$ . However, demanding that the average value of  $\xi_2[\mathbf{x}^{(k)}](s)$  on large scales,  $\bar{\xi}_2[\mathbf{x}^{(k)}]$ , vanishes could be an alternative way to fix  $\beta$ . We also discussed briefly the possibility of using a hybrid approach to break the degeneracy in  $\beta$  and determine both  $f$  and  $b$ . In this sense, the reconstruction can simultaneously give us the cosmological parameter values, the real-space galaxy catalogue, the large-scale peculiar velocity field, and the initial matter density field. Alternatively, one can use the standard methods to determine  $f$ ,  $b$ , and use such values to do reconstruction, which still gives the other quantities and this has the advantage that these quantities are for our particular realization of the universe.

The removal of peculiar-velocity-induced modifications to the real galaxy coordinates means that it is possible to get rid of the need to model the effect of RSD on observables, as we have seen above with the example of galaxy two-point correlation function. The latter is indeed one of the most well-understood probes of galaxy clustering,

<sup>7</sup>As we discussed, in practice only E3, the quadrupole of the reconstructed galaxy catalogue, can be directly used to test convergence as the other estimators require knowledge that is not readily available from observations. However, as we have seen, if the convergence happens for E3, then it happens for the other estimators as well.



**Table 2.** The correlation coefficients before and after the reconstruction procedure. We present the correlation between galaxy and matter distribution, and their combinations. The correlations are characterized in terms of  $k_{90}$  and  $k_{50}$ , which are the  $k$  values where the correlation coefficient falls to 90 per cent and 50 per cent, respectively.  $k_{50}$  and  $k_{90}$  are quoted in unit of  $h \text{ Mpc}^{-1}$ . The results shown here are found by reading the corresponding values from Fig. 8.

Correlation coefficient between	Notation	$k_{90}$	$k_{50}$
<b>Real-space galaxy distribution and</b>			
Redshift-space galaxy distribution	$r[\delta_g^s, \delta_g^r]$	0.23	0.68
Reconstructed real-space galaxy distribution (starting from redshift space)	$r[\delta_g^{(3)}, \delta_g^r]$	0.32	1.02
<b>Initial matter density field and</b>			
Real-space galaxy distribution	$r[\delta_g^r, \delta_{\text{ini}}]$	0.08	0.19
Redshift-space galaxy distribution	$r[\delta_g^s, \delta_{\text{ini}}]$	0.07	0.17
Reconstructed linear matter density from real-space galaxy distribution	$r[\delta_{\text{rec}}, \delta_{\text{ini}}]$	0.12	0.35
Reconstructed linear matter density from redshift-space galaxy distribution (without iterative RSD removal)	$r[\delta_{\text{rec}}^{(0)}, \delta_{\text{ini}}]$	0.11	0.30
Reconstructed linear matter density from redshift-space galaxy distribution (with iterative RSD removal)	$r[\delta_{\text{rec}}^{(3)}, \delta_{\text{ini}}]$	0.12	0.31

and there are other probes for which the modelling of RSD effects is less widely studied. Some examples are the higher order correlation functions of galaxies (e.g. Slepian & Eisenstein 2017) and the cross-correlation of galaxies with clusters (e.g. Zu & Weinberg 2013) or voids (e.g. Hamaus et al. 2015; Cai et al. 2016; Nadathur & Percival 2019). These are situations where reconstruction can be useful by producing reconstructed galaxy catalogues from which the corresponding real-space quantities can be measured; this also has the advantage that the same reconstructed galaxy catalogue can be used to study different probes, rather than having to build a RSD model for each of them. We note that reconstruction-based method has recently been applied to study voids by Nadathur, Carter & Percival (2019). We also note that, even for the redshift-space two-point galaxy correlation function, obtaining per cent-level accuracy in the theoretical predictions down to  $s \sim 20 h^{-1} \text{ Mpc}$  is challenging (see e.g. Bianchi, Chiesa & Guzzo 2015; Bianchi, Percival & Bel 2016; Kuruvilla & Porciani 2018, for some recent efforts to improve the modelling).

Another field where reconstruction can be helpful is the study of dark energy and modified gravity models (see e.g. Baker et al. 2019; Li et al. 2019 for some latest reviews). In these models, large-scale structure formation can be more complicated than in  $\Lambda\text{CDM}$ , making it more challenging to predict RSD effects (e.g. Bose & Koyama 2016; Valogiannis, Bean & Aviles 2019). The removal of such effects through reconstruction can help to simplify the development of theoretical templates for constraining these models (e.g. Koyama, Taruya & Hiramatsu 2009; Aviles & Cervantes-Cota 2017). However, we caution that in the current implementation of our reconstruction algorithm we have used the assumption that linear bias is scale-independent on large scales. Whilst this assumption is a good one for some models, it may break down for others for which a more detailed study is needed to understand how best to incorporate their non-trivial bias behaviour. We note that including galaxy biases beyond linear order has been studied in Birkin et al. (2019) that demonstrated its feasibility.

Finally, in this work we have studied the reconstruction in ‘idealized’ galaxy catalogues in periodic cubic simulation boxes, whilst observational systematics such as survey geometry, selection function, and redshift failure are not included in the analysis. These must be tested carefully using more realistic mock catalogues where such systematics exist (e.g. Smith et al. 2017), such as done by Hada & Eisenstein (2019). One advantage of reconstruction in real space is that survey boundary effects are localized and the induced

error is primarily restricted to regions near the boundaries (Mao & Wang 2019), and we expect the same will apply to our reconstruction algorithm as well. We also expect the effect of redshift evolution on galaxy clustering to be small in our reconstruction method, as long as the redshift dependence of galaxy bias is properly accounted for. Another complication not taken into account in the analysis here is that, even without peculiar velocities, translating a galaxy’s redshift to its radial coordinate requires the assumption of a fiducial cosmological model – an incorrect one would lead to anisotropies in the galaxy correlation function even before RSD effects are included, and these anisotropies could be confused as an RSD signal, leading to incorrect reconstruction outcomes. In the standard approach, it is possible to transform the theoretical prediction of the redshift space correlation function to the fiducial cosmology (e.g. Sanchez et al. 2017) used to infer the correlation function from galaxy redshifts. We expect the hybrid approach mentioned in the end of Section 4 (for which a cosmological model also has to be assumed) will help to break this degeneracy in a similar way, but a detailed investigation of this will be left for future work.

## ACKNOWLEDGEMENTS

We thank Shaun Cole and Jie Wang for useful comments, and Jiamin Hou for early discussions on this subject. BL is supported by the European Research Council via an ERC Starting Grant (ERC-StG-716532-PUNCA) and the UK Science and Technology Facilities Council (STFC) via Consolidated grant no. ST/L00075X/1. MC is supported by the EU Horizon 2020 Research and Innovation programme under a Marie Skłodowska-Curie grant agreement 794474 (DancingGalaxies). This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This equipment was funded by BIS National E-infrastructure capital grant no. ST/K00042X/1, STFC capital grant nos ST/H008519/1, ST/K00087X/1, STFC DiRAC Operations grant no. ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Ade P. A. R. et al., 2014, *A&A*, 571, A24
- Ade P. A. R. et al., 2016, *A&A*, 594, A17
- Alonso D., 2012, preprint ([arXiv:1210.1833](https://arxiv.org/abs/1210.1833))
- Aviles A., Cervantes-Cota J. L., 2017, *Phys. Rev. D*, 96, 123526
- Baker T. et al., 2019, preprint ([arXiv:1908.0343](https://arxiv.org/abs/1908.0343))
- Bianchi D., Chiesa M., Guzzo L., 2015, *MNRAS*, 446, 75
- Bianchi D., Percival W., Bel J., 2016, *MNRAS*, 463, 3783
- Birkin J., Li B., Cautun M., Shi Y., 2019, *MNRAS*, 483, 5267
- Bos E. G. P., Kitaura F.-S., van de Weygaert R., 2019, *MNRAS*, 488, 2573
- Bose S., Koyama K., 2016, *JCAP*, 1608, 032
- Brenier Y., Frisch U., Hénon M., Loeper G., Matarrese S., Mohayaee R., Sobolevskii A., 2003, *MNRAS*, 346, 501
- Burden A., Percival W. J., Manera M., Cuesta A. J., Magana M. V., Ho S., 2014, *MNRAS*, 445, 3152
- Burden A., Percival W. J., Howlett C., 2015, *MNRAS*, 453, 456
- Cai Y.-C., Taylor A., Peacock J. A., Padilla N., 2016, *MNRAS*, 462, 2465
- Cautun M. C., van de Weygaert R., 2011, *Astrophysics Source Code Library*, record ascl:1105.003
- Cautun M., Paillas E., Cai Y.-C., Bose S., Armijo J., Li B., Padilla N., 2018, *MNRAS*, 476, 3195
- Cole S., Fisher K. B., Weinberg D. H., 1994, *MNRAS*, 267, 785
- Crocce M., Scoccimarro R., 2008, *Phys. Rev. D*, 77, 023533
- Croft R. A. C., Gaztanaga E., 1997, *MNRAS*, 285, 793
- Cuesta-Lazaro C., Li B., Eggemeier A., Zarrouk P., Baugh C. M., Nishimichi T., Takada M., 2020, preprint ([arXiv:2002.02683](https://arxiv.org/abs/2002.02683))
- Davis M., Strauss M. A., Yahil A., 1991, *ApJ*, 372, 394
- DESI Collaboration et al., 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
- Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
- Eisenstein D. J. et al., 2005, *ApJ*, 633, 560
- Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, *ApJ*, 664, 675
- Frisch U., Matarrese S., Mohayaee R., Sobolevski A., 2002, *Nature*, 417, 260
- Gramann M., Cen R., Gott J., Richard I., 1994, *ApJ*, 425, 382
- Hada R., Eisenstein D. J., 2018, *MNRAS*, 478, 1866
- Hada R., Eisenstein D. J., 2019, *MNRAS*, 482, 5685
- Hamaus N., Sutter P. M., Lavaux G., Wandelt B. D., 2015, *JCAP*, 1511, 036
- Hamilton A. J. S., 1993, *ApJ*, 406, L47
- Hawkins E. et al., 2003, *MNRAS*, 346, 78
- Hernández-Aguayo C., Hou J., Li B., Baugh C. M., Sánchez A. G., 2019, *MNRAS*, 485, 2194
- Jasche J., Wandelt B. D., 2013, *MNRAS*, 432, 894
- Johnston S. et al., 2008, *Exp. Astron.*, 22, 151
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kitaura F. S., EnBlin T. A., 2008, *MNRAS*, 389, 497
- Kitaura F.-S., Ata M., Rodriguez-Torres S. A., Hernandez-Sanchez M., Balaguera-Antolinez A., Yepes G., 2019, preprint ([arXiv:1911.00284](https://arxiv.org/abs/1911.00284))
- Koyama K., Taruya A., Hiramatsu T., 2009, *Phys. Rev. D*, 79, 123512
- Kuruvilla J., Porciani C., 2018, *MNRAS*, 479, 2256
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Lavaux G., 2016, *MNRAS*, 457, 172
- Li B., 2018, *Simulating Large-Scale Structure for Models of Cosmic Acceleration*, 2514-3433. IOP Publishing, Bristol, UK
- Li B., Koyama K. et al., 2019, *Modified Gravity: Progresses and Outlook of Theories, Numerical Techniques and Observational Tests*. World Scientific, Singapore
- Manera M. et al., 2013, *MNRAS*, 428, 1036
- Mao T., Wang J., Li B., Cai Y., Falck B., Neyrinck M., Szalay A., 2020, preprint ([arXiv:2002.10218](https://arxiv.org/abs/2002.10218))
- Mohayaee R., Frisch U., Matarrese S., Sobolevskii A., 2003, *A&A*, 406, 393
- Monaco P., Efsthathiou G., 1999, *MNRAS*, 308, 763
- Nadathur S., Percival W. J., 2019, *MNRAS*, 483, 3472
- Nadathur S., Carter P., Percival W., 2019, *MNRAS*, 482, 2459
- Nusser A., Davis M., 1994, *ApJ*, 421, 1
- Obuljen A., Villaescusa-Navarro F., Castorina E., Viel M., 2017, *JCAP*, 2017, 012
- Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, *MNRAS*, 427, 2132
- Peacock J. A. et al., 2001, *Nature*, 410, 169
- Peebles P. J. E., 1989, *ApJ*, 344, L53
- Planck Collaboration IX, 2019, preprint ([arXiv:1905.05697](https://arxiv.org/abs/1905.05697))
- Sanchez A. G. et al., 2017, *MNRAS*, 464, 1640
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Schmittfull M., Baldauf T., Zaldarriaga M., 2017, *Phys. Rev. D*, 96, 023505
- Shi Y., Cautun M., Li B., 2018, *Phys. Rev. D*, 97, 023505
- Slepian Z., Eisenstein D. J., 2017, *MNRAS*, 469, 2059
- Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, *MNRAS*, 470, 4646
- Strauss M. A., 1989, PhD thesis. California Univ., Berkeley
- Taylor A. N., Rowan-Robinson M., 1993, *MNRAS*, 265, 809
- Teyssier R., 2002, *A&A*, 385, 337
- Valogiannis G., Bean R., Aviles A., 2020, *JCAP*, 2020, 55
- van de Weygaert R., Schaap W., 2009, *The Cosmic Web: Geometric Analysis*. Springer, Berlin, Heidelberg, p. 291
- Wang X., Pen U.-L., 2019, *ApJ*, 870, 116
- Wang H., Mo H. J., Yang X., van den Bosch F. C., 2012, *MNRAS*, 420, 1809
- Wang H., Mo H. J., Yang X., Jing Y. P., Lin W. P., 2014, *ApJ*, 794, 94
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87
- Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, *ApJ*, 372, 380
- Yu Y., Zhu H.-M., 2019, *ApJ*, 887, 265
- Zhu H.-M., Yu Y., Pen U.-L., Chen X., Yu H.-R., 2017, *Phys. Rev. D*, 96, 123502
- Zhu H.-M., Yu Y., Pen U.-L., 2018, *Phys. Rev. D*, 97, 043502
- Zhu H.-M., White M., Ferraro S., Schaan E., 2019, *MNRAS*, 494, 4244
- Zu Y., Weinberg D. H., 2013, *MNRAS*, 431, 3319

This paper has been typeset from a  $\mathrm{T}_{\mathrm{E}}\mathrm{X}/\mathrm{L}^{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$  file prepared by the author.